

Llicenciatura en Ciències i Tècniques Estadístiques

Títol: *Relació del pes al néixer amb l'exposició al diòxid de nitrogen (NO_2) durant el període d'embaràs*

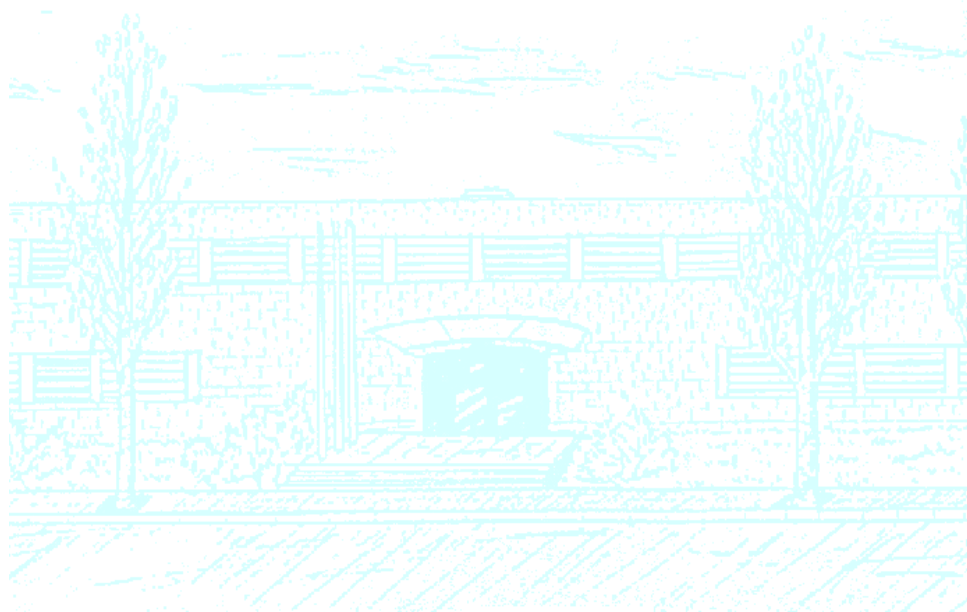
Autor: David Martínez Murciano

Director: Xavier Basagaña Flores

Ponent: Josep Ginebra Molins

Departament: Estadística i Investigació Operativa

Convocatòria: Juny 2010



Facultat de Matemàtiques
i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya

Projecte fi de carrera

Relació del pes al néixer amb l'exposició al diòxid de nitrogen (NO_2) durant el període d'embaràs

David Martínez

9 de Juny de 2010

Director: Xavier Basagaña
Ponent: Josep Ginebra

Índex

Índex.....	4
Índex de taules i figures.....	6
Presentació.....	8
1.1. Descripció del centre.....	8
1.2. Tasques desenvolupades.....	10
1.3. El per què d'aquest projecte final de carrera.....	11
Estudis de contaminació ambiental i salut.....	12
2.1. Introducció.....	12
2.2. Estudis “tipus” de contaminació ambiental.....	13
2.3. Estimació de l'exposició a la contaminació atmosfèrica.....	14
2.4. Mètodes de Mesura i modelatge de la contaminació atmosfèrica.....	15
2.5. Validesa del model de predicció – Cross Validation.....	19
2.6. Error de Mesura – “Errors in variables”.....	21
2.7. Mètodes de correcció de l'error de mesura.....	22
2.7.1. Mètodes Bayesians.....	23
2.7.2. Regression calibration.....	23
2.7.3. Bootstrap.....	24
Justificació de l'estudi.....	26
3.1. Origen de l'Estudi.....	26
3.2. Desenvolupament de l'Estudi.....	27
3.2.1. Resultats obtinguts.....	28
3.2.2. Vincle entre contaminació atmosfèrica i Pes al néixer.....	30
3.2.3. Importància dels estudis epidemiològics.....	31
Objectiu d'aquest projecte.....	32
Recollida de dades de l'estudi.....	34
5.1. Dades de NO ₂	35
5.1.1. Mostreig a 57 punts de Sabadell.....	35
5.1.2. Dades de l'estació de trànsit.....	37
5.2. Dades de Trànsit i geogràfiques de Sabadell.....	37
5.3. Dades de les dones de l'estudi i del nadó al néixer.....	38
Metodologia.....	39
6.1. Model lineal.....	41
6.2. Bootstrap.....	42
Descriptiva de les dades.....	47
7.1. Dades NO ₂	47
7.1.1. Mesures de NO ₂ preses a 57 punts de Sabadell.....	47
7.1.2. Avaluació de les variacions temporals de l'NO ₂ recollit a l'estació de trànsit.....	50
7.2. Descriptiva univariant de les variables d'interès.....	56
7.2.1. Dades de trànsit i geogràfiques per modelar l'NO ₂	56
7.2.2. Covariables de les dones i del nadó.....	58
7.3. Descriptiva bivariant de les variables explicatives i el pes del nadó al néixer.....	61
Resultats dels models.....	64
8.1. Model lineal.....	64
8.1.1. Modelatge i predicció del NO ₂	64

8.1.2.	Descriptiva de les prediccions de l'exposició al NO ₂ respecte la resta de covariables	67
8.1.3.	Model final pel pes al néixer	70
8.2.	Bootstrap	72
8.2.1.	Modelatge i predicció del NO ₂	72
8.2.2.	Model del pes al néixer.....	72
	Conclusions	76
	Limitacions / Treball futur.....	79
	Bibliografia	81
	Annexes.....	84
12.1.	Anexxe 1: Exemple de Qüestionaris	84
12.2.	Anexxe 2: Sintaxi en R de la simulació del bootstrap	101
12.3.	Anexxe 3: Variables de Trànsit.....	104

Índex de taules i figures

Figura 1. Procés entre la formació del contaminant i l'efecte sobre l'individu	14
Figura 2. Exemple d'Error de mesura en una variable explicativa: gràfic de la ingesta de calories mesurades sense error (esquerre) i gràfic de la ingesta de calories mesurada amb error (dreta) respecta el pes	22
Taula 1. Reclutament per cohort.....	27
Taula 2. Seguiment per cohort.....	28
Taula 3. Diferents fonts d'informació del NO ₂ i període de recollida	35
Figura 3. Punts mostrejats de la ciutat de Sabadell.....	37
Figura 4. Diagrama de l'ajust dels models de predicció de l'NO ₂ , i del model pes al néixer incorporant la incertesa de la predicció.....	40
Figura 5. Bootstrap de les bases de dades amb les mesures preses als 57 punts de la ciutat de Sabadell i amb les dades de les dones i del nadó per poder tenir en compte la incertesa de la predicció de l'NO ₂ en el model del pes	43
Taula 4. Comparació de les estimacions obtingudes pel model lineal i el model lineal més bootstrap de les dades simulades en funció de si l'NO ₂ tenia o no efecte en el pes	44
Figura 6. Gràfic dels intervals de confiança predits a l'escenari 1 pel model lineal amb el valor real de la predicció (dreta) i gràfic dels intervals de confiança predits pel model lineal amb bootstrap amb el valor real de la predicció. En vermell aquells intervals que no contenen l'efecte real.	45
Figura 7. Gràfic dels intervals de confiança predits a l'escenari 2 pel model lineal amb el valor real de la predicció (dreta) i gràfic dels intervals de confiança predits pel model lineal amb bootstrap amb el valor real de la predicció. En vermell aquells intervals que no contenen l'efecte real.	46
Figura 8. Valors del NO ₂ a l'abril en els diferents punts mostrejats en funció de la latitud i la longitud	47
Figura 9. Valors del NO ₂ al juny en els diferents punts mostrejats en funció de la latitud i la longitud	47
Figura 10. Valors del NO ₂ a l'Octubre en els diferents punts mostrejats en funció de la latitud i la longitud	48
Figura 11. Valors del NO ₂ al Març en els diferents punts mostrejats en funció de la latitud i la longitud	48
Taula 5. Promig diari de les mesures de NO ₂ recollides als 57 punts de Sabadell durant cadascuna de les campanyes de mostreig	48
Figura 12. Valors mitjans del NO ₂ a las diferents campanyes mostrejades i mitjana global	49
Taula 6. Correlació de les mesures de NO ₂ recollides als 57 punts de Sabadell a cada campanya	50
Figura 13. Gràfics de dispersió de les mesures mostrejades de NO ₂ en els 57 punts de la ciutat durant les diferents campanyes	50
Taula 7. Mitjana mensual recollida per l'estació de trànsit de l'NO ₂ durant els anys 2001-2007	51
Figura 14. Gràfics temporals de la mesura de NO ₂ del captador de trànsit diàriament, per dia de la setmana, mes i any	52
Taula 8. Models del NO ₂ dels anys 2001-2007 del captador de trànsit en funció del dia de la setmana	52

Taula 9. Models del NO ₂ dels anys 2001-2007 del captador de trànsit en funció de l'any de mesura.....	53
Taula 10. Models del NO ₂ dels anys 2001-2007 del captador de trànsit explicat per les variables dia de la setmana, mes i any	54
Taula 11. Descriptiva del promig de la mesura de NO ₂ de l'estació de trànsit durant els mateixos períodes de les campanyes de recollida dels 57 punts de Sabadell	54
Figura 15. Gràfics de les mesures de NO ₂ del captador de trànsit (línia vermella) i de les mesures preses en els 57 punts de la ciutat de Sabadell	55
Taula 12. Descriptiva del tipus d'àrea en un radi de 500 m al punt de mostreig i alçada respecte el nivell del mar	57
Taula 13. Descriptiva de les variables geogràfica tipus de carrer.....	58
Taula 14. Descriptiva de les característiques demogràfiques i d'hàbits de les dones de l'estudi 1.....	59
Taula 15. Descriptiva de les característiques demogràfiques de les dones de l'estudi 2 ..	60
Taula 16. Descriptiva de les característiques de l'embaràs i del nadó de les dones de l'estudi 2.....	60
Taula 17. Descriptiva de les característiques de l'embaràs i del nadó de les dones de l'estudi 2.....	60
Figura 16. Distribució del pes al néixer dels nadons.....	61
Taula 18. Anàlisi bivariant entre el pes al néixer i la resta de covariables de l'estudi	63
Taula 19. Model de regressió de l'NO ₂ als 57 punts mostrejats de Sabadell.....	65
Figura 17. Anàlisi dels residus del model ajustat al promig de les mesures de NO ₂ a les 4 campanyes	65
Taula 20. Predicció de l'exposició de les dones de Sabadell en el període comprès entre les campanyes de mesura	66
Taula 21. Predicció de l'exposició de les dones de Sabadell durant el període d'embaràs	66
Figura 18. Boxplot de la predicció del NO ₂ per les 611 dones residents a la ciutat de Sabadell (esquerra) i Boxplot de la predicció del NO ₂ durant el període d'embaràs de les dones residents a la ciutat de Sabadell	67
Taula 22. Anàlisi bivariant entre l'NO ₂ i la resta de covariables de l'estudi.....	70
Figura 19. Plot del pes al néixer del nadó i els nivells d'exposició a l'NO ₂ de la mare durant el període d'embaràs.	70
Taula 23. Model lineal del pes al néixer fent servir la predicció del NO ₂ obtinguda a través del LUR model	71
Taula 24. Model lineal del valor de NO ₂ recollit als 57 punts mostrejats de Sabadell mitjançant bootstrap	72
Taula 25. Model lineal del pes al néixer incorporant la incertesa de la predicció del NO ₂ mitjançant bootstrap	74
Figura 20. Histograma de l'estimació dels coeficient del model del pes del nadó al néixer	75

Capítol 1

Presentació

Aquest Projecte Final de Carrera es presenta en la modalitat d'Empresa ja que inclou part del treball que he desenvolupat al Centre de Recerca en Epidemiologia Ambiental (CREAL) en els darrers dos anys. A continuació es fa una breu descripció d'aquest centre, del tipus d'estudis que s'hi fan, i de les tasques que hi he desenvolupat.

1.1. *Descripció del centre*

El CREAL és una institució creada a finals de l'any 2005 amb la missió de promoure i desenvolupar recerca epidemiològica avançada sobre els factors ambientals que afecten a la salut humana, per tal de facilitar la prevenció i el control dels seus efectes perjudicials. El CREAL és una iniciativa de la Generalitat de Catalunya amb la col·laboració del Parc de Salut Mar de Barcelona i la Universitat Pompeu Fabra (UPF). Tot i la seva recent constitució, aquest centre és l'hereu d'una trajectòria en recerca epidemiològica ambiental fomentada inicialment per la Unitat de Recerca Respiratòria i Ambiental (URRA) de l'Institut Municipal d'Investigació Mèdica (IMIM-Hospital del Mar).

L'existència del centre es justifica pels grans reptes que encara afronta la salut pública, tant en els països desenvolupats com en els que es troben en vies de desenvolupament, per protegir les poblacions dels efectes d'agents físics, químics i biològics. La recerca en epidemiologia ambiental és crucial per la presa de decisions per part del sector públic pel que fa a la regulació de determinats contaminants i garantir, per exemple, la qualitat de l'aire i l'aigua, o la seguretat alimentària.

La finalitat del centre és la recerca científica, i per tant el producte final són publicacions en revistes científiques especialitzades. L'any 2009 es van publicar 100 treballs amb la col·laboració d'algun membre del CREAL. La recerca té una finalitat pràctica i està encaminada al desenvolupament de polítiques de protecció de la salut que permetin la disminució de les malalties i les discapacitats socials degudes a exposicions ambientals. A part de la producció científica, el CREAL també proporciona serveis d'alta consultoria i assistència

tècnica a les administracions públiques i entitats privades en aquells problemes de salut ambiental directament relacionats amb la seva experiència investigadora. Així, els investigadors del CREAL ofereixen assessorament realitzant estudis o informes de possibles perills o de recomanacions de salut pública.

Els reptes del CREAL s'aconsegueixen mitjançant projectes de recerca competitiu, educació avançada i col·laboracions multidisciplinàries nacionals i internacionals. El personal investigador que hi treballa té una llarga experiència en recerca i en assessorament de sistemes d'informació, avaluació de riscos ambientals i gestió de situacions de crisi, així com de formació en mètodes d'epidemiologia ambiental i gestió del coneixement.

La recerca del CREAL es centra en 6 programes de recerca: Respiratori, Càncer, Infància, Contaminació atmosfèrica, Contaminació de l'aigua i Radiacions. A més, també es duen a terme altres projectes associats a la recerca realitzada com són els de Bioestadística o els Educacionals. En la majoria de projectes del CREAL s'avalua l'exposició a contaminants ambientals i la relació d'aquesta exposició amb la salut.

A tall d'exemple, alguns dels projectes del CREAL són:

- **HI-Wate:** L'objectiu general d'aquest estudi és el d'investigar els riscos potencials per a la salut humana associats amb l'exposició a llarg termini a baixos nivells de desinfectants (com el clor) i els derivats de desinfectants (DBP) que es creen en l'aigua per ús i consum humà, i els que s'utilitzen en la indústria alimentària. L'estudi inclou l'anàlisi dels riscos i beneficis, inclosa l'avaluació quantitativa del risc associat a la contaminació microbiana per la ingestió d'aigua i el risc químic, així com la comparació d'opcions alternatives de tractament. Els resultats comportaran avenços importants en l'avaluació i gestió del risc.
- **ESCAPE** – “European Study of cohorts for air pollution effects”: ESCAPE és una col·laboració de més de 30 estudis de cohort europeus, que inclou uns 900.000 participants. El seu objectiu és mesurar l'impacte dels contaminants atmosfèrics en la salut i reduir la incertesa que es té tant en la mesura d'aquests contaminants, com de l'efecte que tenen. ESCAPE també posarà a prova noves hipòtesis sobre els efectes específics de salut de la contaminació atmosfèrica. Per desenvolupar aquests objectius, s'avaluaran les diferències en els nivells de contaminació atmosfèrica dins una mateixa ciutat, dins una mateixa zona i dins d'un mateix país amb mètodes avançats d'avaluació de l'exposició. ESCAPE se centrarà en quatre

categories d'estudis de cohorts: 1) Estudis de l'embaràs i de cohorts de nounats; 2) Estudis sobre malalties respiratòries en adults; 3) Estudis sobre malalties cardiovasculars en adults; 4) Estudis sobre la incidència de càncer i mortalitat.

- TAPAS – “Transportation, Air pollution and Physical ActivitieS; an integrated health risk assessment progamme of climate change and urban policies”: Grups internacionals han recomanat canvis en la política de control de l'obesitat, la contaminació i el canvi climàtic. Planificar i promocionar polítiques on es promogui moure's a peu i en bicicleta poden contribuir a aquesta fita, aconseguint tots dos objectius alhora. El fet de continuar incrementant l'exposició al trànsit d'aquelles persones que realitzen modes de transport actius, com són caminar o anar en bicicleta pot tenir conseqüències no desitjades. L'objectiu d'aquest estudi és precisament el d'ajudar a la presa de decisions sobre les polítiques dirigides a lluitar contra el canvi climàtic i millorar la salut dels ciutadans. En particular, a assessorar sobre les condicions i la manera de promoure el transport actius, i per tant, la millora en la salut.

1.2. Tasques desenvolupades

Al juliol del 2008 vaig començar a treballar com a estadístic al Centre de Recerca en Epidemiologia Ambiental (CREAL) sota la direcció dels doctors Mark Nieuwenhuijsen i Martine Vrijheid. Sóc l'estadístic de diversos projectes d'investigació i una de les meves tasques consisteix en fer de data manager en algun dels projectes en que col·laboro. El fet que les dades provenguin de diferents fonts d'informació (per exemple: qüestionaris, aparells de captació, sistema geogràfic d'informació (GIS)) fa que s'hagin d'ordenar, agrupar i depurar. És el cas del projecte Commuters, que pretén estudiar els nivells de contaminació als que estan exposades les persones en funció del tipus de transport que es faci servir per moure's.

Principalment, però, dedico la major part del temps a l'anàlisi de les dades. Les anàlisis més comunes que he dut a terme han estat models logístics, en els que s'estudiava el increment de risc de certes malformacions en nadons (per exemple: hipospadies, que és un defecte del tracte urinari dels nens), o models lineals sobre el desenvolupament neurocognitiu dels nadons (Vrijheid et al.). També he realitzat diversos metanàlisis per tal d'avaluar la relació de diversos compostos orgànics presents a l'aigua i contaminants de l'aire amb certes malformacions congènites (per exemple, defecte septal ventricular, que és un orifici a la paret muscular que separa els ventricles esquerra i dret del cor) a través d'estudis existents. Altres

anàlisi que he desenvolupat durant aquest temps han estat la realització d'un cluster anàlisi per tal de determinar si existia una agrupació entre els nivells d'exposició a un conjunt de contaminants (organoclorats, els policloribifenils – PCB- són un tipus de compostos orgànics amb 1 a 10 àtoms de clor units amb bifenil, i degut a la seva capacitat ininflamable, són presents a transformadors elèctrics, entre d'altres; i mercuri). Per últim, he dedicat bona part de la meua estada al CREAL a desenvolupar un model bayesià que redueixi l'error de mesura en l'estimació de l'exposició a un determinat contaminant, per tal de poder estimar millor l'efecte d'aquest sobre el nadó al néixer. En aquest cas dispo de diverses mesures d'exposició al contaminant, mesures personals, i dins i fora de la llar.

1.3. El per què d'aquest projecte final de carrera

La major part dels estudis en els que participo intenten associar els nivells d'exposició a un contaminant ambiental i algun problema de salut. Normalment l'exposició de cada persona no es coneix i es fan servir diversos mètodes per predir-la, i molts d'aquests mètodes s'acaben basant en models estadístics. No obstant, moltes de les incerteses que es tenen en l'estudi no s'acaben incorporant en aquests models estadístics. L'objectiu d'aquest projecte és tornar a analitzar unes dades d'un estudi de contaminació atmosfèrica intentant incorporar algunes millores en l'anàlisi. Aquestes millores, que permetran introduir en el model algunes de les incerteses que els models més senzills no tenen en compte, s'incorporaran en estudis futurs i seran un primer pas per ajustar models més complexos.

En la següent secció es descriuran dos estudis de contaminació atmosfèrica que podrien ser "l'estudi tipus" en els estudis de contaminació ambiental, i es descriuran les dificultats que presenten aquests estudis i com s'intenten solucionar.

Capítol 2

Estudis de contaminació ambiental i salut

2.1. Introducció

La contaminació atmosfèrica és un problema de salut pública que causa més de 380.000 morts només a la Unió Europea, i més de 800.000 a tot el món, segons dades de World Health Organisation (WHO). Actualment, se sap que la contaminació que prové del trànsit està associada a alguns efectes sobre la salut. Tanmateix, encara queden preguntes per resoldre. És per això que una de les principals àrees d'estudi de l'epidemiologia és l'avaluació de l'exposició ambiental (ja sigui en aigua, terra o aliments).

L'Epidemiologia és la disciplina científica que estudia la distribució, freqüència, determinants, relacions, prediccions i control dels factors relacionats amb la salut i malaltia a poblacions humanes. És una ciència bàsica de la medicina preventiva i una font d'informació per la salut pública. Per tal de desenvolupar els seus objectius, l'epidemiologia es basa en estudis observacionals. És a dir, a diferència d'altres disciplines, on es realitzen estudis experimentals, l'investigador es limita a recollir la informació i a analitzar-la sense intervenir-hi. Tot i això, en molts casos, s'han realitzat estudis previs en animals per confirmar algunes de les hipòtesis definides abans de realitzar els estudis epidemiològics sobre els humans. El realitzar un experiment sobre humans és èticament impossible. No es pot exposar a un grup d'éssers humans a un component que es creu que té components nocives per la salut, és per això que la única alternativa són els estudis observacionals.

Segurament el primer estudi d'epidemiologia en contaminació ambiental es va donar durant l'episodi de boira (smog) de Londres (Bates DV, Bell ML et al.). A partir d'allà es va acceptar el fet que només l'elevada exposició a altes concentracions de contaminació ambiental s'associava amb la mortalitat. Des d'aleshores s'han realitzat molts estudis i publicacions i s'ha demostrat que no només l'exposició a altes concentracions de contaminació ambiental està relacionada amb la mortalitat, sinó que també està associada amb efectes

crònics i greus en diverses malalties, la majoria d'elles respiratòries i cardiovasculars, a baixes concentracions de contaminació.

Trobar efectes de la contaminació per a altes dosis pot ser relativament senzill, però quan es vol demostrar que existeix una associació entre contaminació i problemes de salut a dosis més baixes la cosa es complica. El problema més important és conèixer quina és la dosi real de contaminació a la que està exposada cada persona. Aquí hi intervenen molts factors, com la variabilitat temporal i espacial de la contaminació, on passa la gent la major part del seu temps i determinats hàbits personals. A més a més, la contaminació atmosfèrica té molts components, no tots tenen la mateixa toxicitat, i la barreja de compostos químics pot ser diferent en un lloc o en un altre.

2.2. Estudis “tipus” de contaminació ambiental

L'estudi ARIBA (“Air Pollution and Reproduction in Barcelona”) i l'estudi INMA (Infancia y Medio Ambiente) tenen com a objectiu estudiar l'efecte de diferents contaminants ambientals (diòxid de Nitrogen, NO₂; partícules en suspensió de diàmetre inferior a 10 µg, PM₁₀; i compostos orgànics volàtils (VOCS) en dones embarassades i els seus nounats. L'estudi INMA és l'estudi en el qual s'emmarca aquest projecte i es descriurà posteriorment a la secció 2.1. D'altra banda, l'estudi ARIBA es troba en una fase menys avançada que l'estudi INMA. Aquest projecte té l'objectiu de tornar a analitzar dades de l'estudi INMA incorporant algunes millores en l'anàlisi, i per tal d'incorporar-les als futur anàlisis de les dades de l'estudi ARIBA. A la secció 1.4.4. es fa una descripció de com es realitzen habitualment els estudis dels efectes de la contaminació ambiental en la salut i de quins són els principals problemes amb els que s'enfronten.

L'estudi ARIBA disposa d'una base de dades coordinada per l'Hospital Clínic amb més de 15000 naixements provinents tant d'hospitals públics com privats. En aquesta base de dades, a més de disposar de variables referents al seguiment de l'embaràs i les condicions del part i naixement del nadó, també es disposa d'informació referent al domicili, i hàbits de la dona vers l'alcohol, el tabac i la presa de medicaments. Pel que fa als contaminants ambientals, s'ha recollit informació referent als nivells dels contaminants a diferents punts de la ciutat de Barcelona. A més, actualment s'estan realitzant campanyes de mesura dels contaminants ambientals a les llars d'una submostra de la població en estudi i als seus llocs de treball, així com mesures personals (cada dona porta un captador durant 7 dies). L'objectiu d'aquest estudi serà el de utilitzar les diferents mesures preses de contaminació per obtenir una

estimació individual d'exposició a un contaminant, i així poder estudiar l'efecte del contaminant amb problemes durant l'embaràs, i possibles malformacions del nadó al néixer.

2.3. Estimació de l'exposició a la contaminació atmosfèrica

La contaminació atmosfèrica urbana, que té com a principal font d'emissió el trànsit, no és fàcil de quantificar. La barreja de contaminants a l'aire i la variabilitat espacial i temporal dificulta la seva estimació, i la mobilitat de les persones i els hàbits personals compliquen encara més l'obtenció del nivell d'exposició individual. Un exemple d'aquesta variabilitat és l'òxid de nitrogen (NO_x) que pot tenir més variabilitat dins una mateixa ciutat que entre ciutats (Briggs 2000). És per aquest motiu, i per al increment del trànsit a les ciutats (font principal de la contaminació atmosfèrica en aquestes), que aquest fenomen s'ha convertit en una de les principals àrees d'investigació (Brunekreef y Holgate, 2002). A la figura 1 es pot veure el procés existent entre la formació del contaminant i l'efecte sobre les persones.

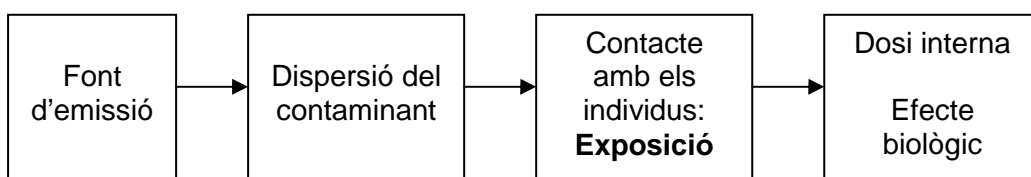


Figura 1. Procés entre la formació del contaminant i l'efecte sobre l'individu

Els principals contaminants produïts pel trànsit a les ciutats europees són el Monòxid de Nitrogen (NO_x), monòxid de carboni (CO), BTEX (benzè, toluè, etilbenzè i xelens), fums negres (BC) i partícules ultrafines. La concentració d'aquestes varia en funció de la proximitat al trànsit. Per exemple, la concentració de NO_x i de partícules ultrafines (origen primari, donat que es forma de manera natural) baixa considerablement a una distància superior als 100 m. En canvi, per l' NO_2 , el gradient en funció de la distància al trànsit no és tan pronunciat. En zones on la dispersió dels contaminants és molt limitada per l'alçada dels edificis la concentració pot ser molt més elevada.

El interès dels estudis epidemiològics, i d'aquest projecte, es centra en obtenir la millor mesura personal d'exposició a la contaminació ambiental i així poder estudiar el seu efecte d'aquesta sobre la salut. S'ha de considerar que degut als problemes logístics i econòmics derivats de la presa de les mesures d'exposició ambiental, existeixen diferents mètodes de recollida de dades i tècniques estadístiques per tal d'obtenir la millor mesura personal possible. Per exemple, les mesures de la concentració ambiental de contaminants es donen

en mitjanes anuals, diàries o horàries, perdent la possible variabilitat temporal, fet que és menys preocupant si el que es vol són exposicions cròniques o concentracions acumulades durant un període llarg de temps. Les guies habitualment recomanen mitjanes anuals i com a mínim mitjanes de 24 hores. La disponibilitat d'aquestes mesures depèn tant del contaminant com del mètode utilitzat per prendre la mesura.

Per tal d'avaluar diferents estimacions de la mesura personal d'exposició ambiental tindrem en compte dos factors: el primer, com es recullen les dades que es faran servir per estimar l'exposició ambiental dels individus; el segon, el mètode estadístic que es fa servir. A continuació, detallarem els mètodes de captació de la mesura i els mètodes estadístics més emprats en epidemiologia. A partir d'aquests dos factors la mesura té millor o pitjor qualitat.

2.4. Mètodes de Mesura i modelatge de la contaminació atmosfèrica

Un primer tipus de mesures d'exposició ambiental, són les mesures preses per *Monitors centrals fixes*: Consisteixen en estacions de monitorització fixes situades a diferents punts de les ciutats, que es classifiquen en estacions de trànsit, indústria o de fons. La categoria depèn de la proximitat a les fonts de contaminació. Habitualment, diverses estacions de diferents categories es troben situades en diversos punts dels llocs urbans, tot i que habitualment la mesura que s'acostuma a fer servir és la de l'estació de fons. La situació, en relació a la població i a la font de contaminació, s'ha de tenir en compte en els estudis epidemiològics, donat que aquesta pot tenir un efecte sobre la qualitat de les estimacions i la comparabilitat dels resultats entre ciutats (de Hartog et al.). Les estacions donen nivells de concentració i no nivells d'exposició. Aquestes estacions són instal·lacions realitzades per altres institucions (com per exemple el servei català de meteorologia) que també tenen interès en aquest tipus de mesures. Es per això s'intenten aprofitar aquestes dades per tal d'abaratir costos, tot i que de vegades, el fet de no poder decidir la ubicació d'aquest instrument no permet el seu ús.

Moltes vegades, en epidemiologia, els valors obtinguts per les estacions són utilitzats com a marcadors de l'exposició a la contaminació ambiental, fent-los servir a nivells de regió, ciutat o barri. Tot i això, aquestes mesures amb l'ajuda de *Models de proximitat*, es fan servir per estimar l'exposició ambiental personal. Els models de proximitat es basen en mesurar la proximitat d'un individu (fent servir la direcció de la seva llar, de la feina entre d'altres) a una font de contaminació (un exemple pot ser la distància a carreteres), o bé, assignar l'exposició mesurada per l'estació fixa més propera. En diversos estudis s'ha demostrat que el fet de viure prop de zones amb elevada contaminació és un factor de risc de patir asma, sibil·làncies i

menor funcionament pulmonar en nens (Gauderman et al., 2005). De totes maneres, al comparar aquests mètodes tant senzills amb d'altres més complexos, s'observa la necessitat de fer servir mètodes més complexos on es puguin tenir en compte altres variables com la heterogeneïtat de les zones de residència (diferents tipus d'edificis) o el tipus de dispersió (si és igual en qualsevol direcció).

Altres mètodes que es poden fer servir amb aquest tipus de dades d'exposició ambiental són els *Models d'interpolació espacial*, que es basen en la creació de mapes de contaminació atmosfèrica a partir d'una sèrie de mesures preses en uns punts determinats fent servir tècniques geostatístiques. Un d'aquests mètodes s'anomena Kriging, i el que fa és fer la predicció de la contaminació d'un punt a partir de la mitjana ponderada dels punts més propers. Aquest mètode té l'avantatge que no necessita variables geogràfiques addicionals, i que la fiabilitat del model es pot quantificar mitjançant l'error estàndard. Per contra, aquest mètode tendeix a suavitzar en excés la variabilitat existent tot i que en menor mesura que els models de proximitat. És per això que tampoc no acaba d'aproximar bé la variabilitat existent dins una ciutat. A més, en cas que no existeixi aquesta component espacial, el model no aporta res.

Un segon mètode de mesures de contaminació ambiental habitual en epidemiologia són les *Mesures individuals*. A diferència de les mesures centrals que es fan servir per objectius de regulació, les mesures individuals es fan servir habitualment per la investigació. Es prenen dos tipus de mesures individuals: mesures a les llars de les persones (a vegades també es prenen mesures al lloc de treball); o mesures personals. Les mesures preses a les llars dels individus així com les mesures personals permeten tenir una mesura individual per a cada subjecte. La mesura personal probablement sigui la millor mesura possible per tal d'avaluar l'exposició individual a la contaminació ambiental, però requereix un dispositiu amb característiques especials com connexió a la llum, o bateries. Les dades de contaminació s'obtenen realitzant mesures durant 1 o 2 setmanes en diverses campanyes durant diversos mesos, per tal d'aconseguir que la mitjana d'aquestes representi bastant fidelment la mitjana anual per a cada punt mostrejat (Lebreil et al. 2000). Aquest fet pot provocar la presència d'estacionalitats en les mesures preses. A més, aquest tipus de mesures tenen un elevat cost, pel que difícilment es porten a terme en grans poblacions (Nieuwenhuijsen MJ).

Per tal de modelar aquesta informació, si es disposa de la situació del lloc de mostreig, es poden aplicar els mètodes d'estimació de l'exposició comentats en el cas anterior. Si a més d'aquesta informació també es disposa d'informació referent al trànsit que envolta el punt de mostreig i a altres característiques geogràfiques de la zona, es pot fer servir el que en epidemiologia s'anomena *Land use regression model* o *mapes de regressió* (LUR models).

Aquests models són els més actuals fets servir fins ara tant a ciutats d'Europa com de Nord-Amèrica. Tenen una resolució espacial més alta que la resta de models i per tant, són més acurats a l'hora d'avaluar la variabilitat intraurbana de la contaminació atmosfèrica (Hoek et al., 2008). Aquest mètode va ser desenvolupat per Briggs et al (1997). Fent servir els nivells de contaminació mesurats en uns quants punts de mostreig i una sèrie de variables obtingudes a partir d'un Sistema d'Informació Geogràfica (SIG), com ara dades de trànsit, tipus de zona (rural, urbana, industrial), o densitat de població entre d'altres, es construeix un model de regressió. Un cop obtingut el model de regressió, es realitza la predicció per aquells punts no mostrejats dels quals es disposa de la informació de les variables explicatives fetes servir al model. Amb aquesta darrera metodologia s'acostuma a obtenir, amb un petit nombre de variables, un coeficient de determinació elevat ($R^2 > 0.65-0.70$), fins i tot en aquells contaminants amb gran variabilitat a petita escala (NO_x).

Els desavantatges d'aquest mètode són, principalment, l'especificitat per l'àrea en estudi (no és extrapolable a altres àrees), i la necessitat de dades procedents de SIG d'alta definició, així com un nombre de punts de mostreig suficient i representatiu de l'àrea d'estudi. Una altra limitació important d'aquest mètode és que no té en compte la variabilitat temporal, pel que, en casos com el nostre en el que volem estudiar l'exposició durant un període determinat s'han de buscar altres ajustos addicionals. Una possible solució és estimar la tendència temporal mitjançant els valors registrats diàriament en una estació fixa durant el període d'interès, assumir que el patró temporal és el mateix a tots els punts de la ciutat, i fer les correccions pertinents. A més, quan es té una àrea força heterogènia, és a dir, amb presència de zones rurals i urbanes, una possible solució per tal d'evitar la limitació dels LUR podria ser el combinar-los amb la metodologia del Kriging (Íñiguez et al., 2009).

Per últim tenim els *qüestionaris i modelatge*. Els qüestionaris són una alternativa per tal d'obtenir la mesura d'exposició ambiental individual autoreportada. Aquests han de recollir informació sobre el trànsit a prop de la llar, així com una descripció detallada de les activitats que realitza. La combinació de la informació recollida pot donar-nos una estimació de l'exposició. Tot i això, el fet que la informació sigui recollida en qüestionaris pot donar lloc a diversos biaixos (Hunter et al.) (per exemple es pot reportar menor freqüència d'assistència a bars on l'exposició és més alta que en altres ambients). El modelatge és una eina relativament nova i prometedora d'avaluació de l'exposició ambiental gràcies a l'expansió del Sistema d'Informació Geogràfica (GIS) en el món de l'anàlisi de l'exposició. En aquest cas es fa servir directament la informació recollida en el qüestionari, així com mesures de contaminació al

voltant de la zona de residència. A partir del model més convenient en funció de la informació reportada, es realitza la predicció per la resta de la població.

Existeix un altre tipus de model per estimar l'exposició ambiental personal, que són els *models de dispersió*, però que no són aplicables a cap dels casos anteriors. Són models dinàmics de processos de dispersió que es basen en la ploma de Gauss (aquesta assumeix que les concentracions del contaminant, tant vertical com horitzontal, poden representar-se mitjançant corbes basades en la distribució normal) i que necessita dades de fonts d'emissió, concentracions emeses, dels contaminants, meteorològiques i topogràfiques, les quals es calibren fent servir dades d'estacions fixes. Aquests models tenen com a principal avantatge que a més de capturar la variabilitat espacial, també tenen en compte la variabilitat temporal. El principal problema és l'elevat cost d'obtenir aquestes dades així com l'elevat volum de dades necessàries, tot i que en algun cas s'han fet servir (Nyberg et al., 2000).

En tots els casos descrits es disposa de dades de contaminació ambiental amb una component espacial, així com l'adreça dels participants de l'estudi. A partir d'aquí es busca quin és el millor mètode per poder predir el nivell d'exposició ambiental dels individus basant-se en la seva adreça. S'ha de tenir en compte però, que en cap dels casos anteriorment detallats es té en compte que no tothom es passa el dia a casa, ni tan sols, que sigui el lloc on més temps passa del dia. Aquesta és una part important d'error de la nostra mesura, però no serà objectiu d'aquest projecte, per tant, no hi entrarem al detall. Tot i això, en cas que es pogués disposar de mesures dins i fora de la llar i del lloc de treball i de les activitats que es realitzen durant el dia i el lloc es poden intentar incorporar aquestes dades en models de predicció més complexos.

El que sí que s'estudiarà en el projecte és el fet que les nostres mesures d'exposició a contaminació siguin prediccions de models i no mesures reals, i que per tant les prediccions tenen un error de predicció associat que cal tenir en compte quan es fan servir aquestes prediccions com a variable explicativa en un nou model.

Així doncs, ens trobem amb dos problemes ben diferenciats. El primer problema és com trobar el millor mètode de predicció del NO_2 durant el període d'embaràs amb la informació disponible; el segon, com incorporar l'error degut al fet que treballem amb una predicció, i per tant, una mesura amb error, enlloc de la mesura real.

Sens dubte, a més detall de les mesures preses, més bo podrà ser el model de predicció que s'obtindrà, i per tant, menys error tindrem a la nostra predicció. Però com moltes vegades

això no serà controlable, ens hem de fixar en el model utilitzat per realitzar la predicció. En molts casos, per avaluar la bondat del model de predicció es fa servir el coeficient de determinació (R^2) del model tot i que fer servir aquest mètode dona lloc a la infraestimació de l'error real de predicció. El problema ve donat pel fet que s'està fent servir la mateixa base de dades que s'ha utilitzat per estimar els paràmetres per validar-los. Per tal d'estimar com de bo és el nostre model de predicció farem servir el cross validation.

2.5. Validesa del model de predicció – Cross Validation

El cross validation és un mètode que, basant-se en l'estimació de l'error de predicció, permet avaluar l'ajust d'un model predictiu a unes dades independents.

Per avaluar un model lineal s'acostuma a fer servir el coeficient de determinació, R^2 , basat en el mètode dels residus. El mètode dels residus però, no aporta informació sobre com de bé ajustaria el model a unes noves dades independents. Mitjançant el cross validation es pot incorporar aquest fet en el càlcul de la R^2 .

Per veure la diferència entre un i altre mètode veiem primer com es calcula el coeficient de determinació en el cas del mètode dels residus. Aquest mètode es basa en el percentatge de variabilitat explicada dividit per la variabilitat total:

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$$

essent el numerador la suma de quadrats dels residus i el denominador la suma de quadrats total.

El coeficient de determinació ajustat és igual a l' R^2 però ponderant pel número de variables que s'inclouen en el model:

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \cdot \frac{n - 1}{n - p - 1}$$

Tot i aquesta correcció, aquest coeficient de determinació fa referència al percentatge de variabilitat explicada del model real, però no pel model ajustat.

El mètode de cross validation soluciona, a partir d'una mostra de dades externa o una submostra de les dades originals, el problema del coeficient de determinació. El que fa és realitzar la predicció de la variable que es vol predir sobre aquestes noves dades i la compara amb els nous valors observats. Si el model de predicció és bo els nous valors observats haurien de ser propers a les prediccions \hat{V}_i . Per avaluar aquesta proximitat es fa servir l'error mitjà de predicció, definit com:

$$MSPE = \frac{\sum E[(Z_i - \hat{V}_i)^2]}{n} = \frac{\sum Var(Z_i) + V(\hat{V}_i)}{n} = \frac{\sigma_z(n+p+1)}{n}$$

De manera que l' R^2 proposat quedaria determinat per:

$$R_{FFE}^2 = 1 - \frac{\sigma_z(n+p+1)}{\sigma_y(n+1)}$$

Essent el numerador l'estimació de la suma de quadrats del error de predicció i el denominador la suma de quadrats total del model.

En el nostre cas hem fet servir el mètode de cross validation 10-fold. Aquest mètode es basa en subdividir les dades en 10 mostres diferents. Aleshores, s'agafen 9 d'aquestes 10 mostres per ajustar el model del NO_2 . Amb les estimacions d'aquest model es realitzen prediccions sobre la 10a mostra, que no s'ha fet servir en l'ajust del model. Això es repeteix per cadascuna de les 10 mostres. D'aquesta manera s'obté una predicció per cadascuna de les observacions de la mostra total, però les prediccions s'han obtingut com si es tingués una mostra externa i independent de les dades. Aleshores, aplicant l'equació anterior obtenim una R^2 pel model de predicció que tingui en compte precisament aquest fet.

El segon problema amb el que ens trobarem és que, en el segon model, quan tractem l' NO_2 com una variable explicativa en el model del pes del nadó, la mesura de NO_2 que fem servir no és una mesura real, sinó una predicció amb els seu corresponent error de predicció associat. Aquest fet provoca que, a diferència del que passa en altres àmbits de la investigació (més experimentals), treballem amb una variable explicativa mesurada amb error. A continuació comentarem el problema que això comporta i possibles solucions.

2.6. Error de Mesura – “Errors in variables”

Els models de regressió habituals assumeixen que els predictors estan mesurats sense error. Quan ens trobem amb aquesta situació es produeix el que s'anomena en epidemiologia error de mesura, o “errors invariables” en la literatura estadística. En epidemiologia ambiental, aquest error es dona quasi sempre i en diferents nivells. Per exemple, quan es fa servir la mesura d'una setmana com a exposició mitjana en l'últim any, la mesura que fem servir serà amb tota seguretat com a mínim una mica diferent del valor anual real, i per tant la variable explicativa tindrà un error de mesura. Però fins i tot si el que mesurem està mesurat amb exactitud, però només es fan mesures en una submostra de la població (com és habitual en els estudis epidemiològics com s'ha descrit en la secció 1.4.3) i per la resta es fan servir prediccions, aquestes estimacions són també una font d'error de mesura. Aquest últim és el tipus d'error de mesura del qual intentarem corregir els seus efectes en aquest projecte.

Els errors de mesura que es donen es poden considerar sistemàtics o aleatoris. Un error és sistemàtic quan es pot definir com la diferència entre la mitjana dels valors mesurats menys el valor real. Això passa, per exemple, quan un procés de calibració està basat en un estàndard erroni. L'efecte que té aquest tipus d'error sobre la relació exposició-efecte no és crític donat que no comporta biaix en els efectes de la variable; únicament desplaça la recta de regressió per sobre o per sota però el pendent de la recta és el mateix. A més, en cas que es conegui l'error, la seva correcció és molt senzilla. Per contra, l'error aleatori és més perillós donat que provoca biaix entre la relació de l'exposició i l'efecte. Aquest tipus d'error es pot donar a qualsevol nivell del procés de mesura (ex: variació del volum d'aire que rep el captador, variació en el temps d'exposició). L'efecte d'aquest tipus d'error no es pot predir a priori sense fer determinades assumpcions, i l'única manera de quantificar-ho és a partir de mesures repetides.

Tot i això, aquest tipus d'error es pot donar de manera no diferencial, és a dir, que l'error de la mesura no tingui relació amb el valor de la variable resposta, o de manera diferencial, quan l'error de mesura varia segons els valors que pren la resposta. Un exemple d'error de mesura diferencial és als qüestionaris de dieta, quan s'estima la ingesta de calories. Les persones amb sobrepès acostumen a reportar menys calories de les que acostumen a consumir, i en canvi, les persones amb problemes de baix pes, acostumen a reportar més calories de les que consumeixen. És a dir, que en funció del pes de la persona, l'error a l'estimació de la mesura varia. En canvi, si tenim una bàscula que està mal calibrada, aquesta, independentment del pes, sempre reportarà el mateix error. En el nostre cas, creiem que l'error amb el que ens trobarem serà no diferencial donat que en el nostre cas, l'error vindria a

ser més semblant al de la “bàscula mal calibrada” que no pas a l’error a l’exemple de la ingesta de calories.

Un exemple hipotètic gràfic del que pot comportar l’error de mesura el podem observar a la figura 2. Els gràfic de l’esquerra representa la ingesta de calories mesurades a partir del qüestionari de dieta, i per tant, amb error, i el gràfic de l’esquerra representa la mesura pressa a partir d’un biomarcador, i per tant, sense error. Si comparem les dues rectes de regressió s’observa com en el cas de la mesura amb error existeix una atenuació de l’efecte de la ingesta de calories respecte la recta obtinguda d’ajustar la mesura sense error. A més, s’observa l’existència de sobredispersió respecte a la mesura real.

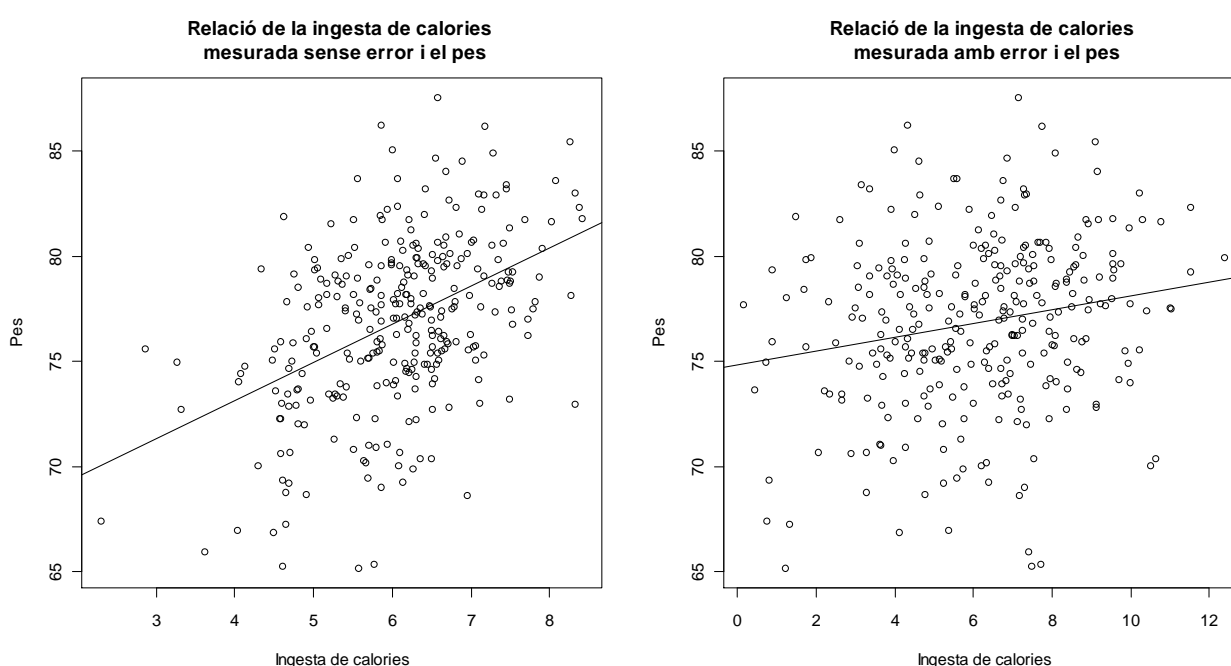


Figura 2. Exemple d’Error de mesura en una variable explicativa: gràfic de la ingesta de calories mesurades sense error (esquerre) i gràfic de la ingesta de calories mesurada amb error (dreta) respecta el pes

Per tal de corregir aquest tipus d’error diverses tècniques han estat utilitzades a la pràctica habitual en epidemiologia.

2.7. Mètodes de correcció de l’error de mesura

Alguns dels mètodes emprats en epidemiologia com a correctors de l’error de mesura són els models bayesians, regression calibration o el bootstrap. A continuació realitzem una petita descripció de cadascun d’ells.

2.7.1. Mètodes Bayesians

El mètode bayesià permet assignar distribucions als predictors, i d'aquesta manera tenir en compte l'error de mesura. A més, es poden ajustar dos models de forma simultània, com és el cas en aquest projecte, on hi ha un model de predicció de l'exposició al diòxid de carboni (a partir de les mesures presses a diferents punts de la ciutat de Sabadell es realitzarà la predicció de l'exposició per les dones participants a l'estudi), i la incertesa de les prediccions del contaminant es transmeten al segon model (model del pes del nadó al néixer respecte l'exposició al NO₂ durant el període d'embaràs). D'aquesta manera, l'efecte estimat en el model resultant tindrà en compte el fet que la mesura té error. Aquesta tècnica s'ha aplicat, per exemple, a estudis on es vincula l'exposició al tabac i el número de cigarretes fumades (Gustafson P. et al.), o en models de l'efecte de la contaminació ambiental sobre la capacitat pulmonar (Molitor et al.). En aquest últim article es realitza un model bayesià multinivell donat que les mesures són preses en diferents ciutats.

2.7.2. Regression calibration

Aquesta metodologia només es fa servir en aquells casos en els quals es disposa de la mesura real (X) per una submostra de la població en estudi. És a dir, quan es disposa de dos jocs de dades. Un primer, amb la informació referent al total d'individus en estudi, on s'inclou la variable de resposta, una covariable Z, que es un "surrogate" de la mesura d'exposició real, i altres covariables (Z). Aleshores, la tècnica de "regression calibration" es basa en el reemplaçament de la variable no observada pel total de la població (X), per la regressió de X sobre altres covariables mesurades (Z,W) sobre part de la població a la qual s'ha observat aquesta. Després de realitzar aquest ajust es realitza l'anàlisi habitual. En aquest cas, a diferència del mètode bayesià, són necessàries unes dades extres per tal de poder desenvolupar-ho. Moltes vegades, aquestes dades són medicions dels contaminant que s'han fet de manera extraordinària a una petita mostra del total de la població, i que recullen de manera més fidel els valors de exposició a la contaminació. En epidemiologia, quan es disposa d'unes dades que recullen el valor real de l'exposició ambiental, o la mesura que es considera més adequada, se l'anomena gold standard. Per exemple, quan es fan qüestionaris de dieta, és l'habitual l'ús de les analítiques de sang per corregir el possible error introduït al qüestionari de dieta. Tot i això, d'aquesta manera només es corregeix l'error present a l'estimar l'efecte d'aquesta amb la variable resposta, però no corregeix la desviació estàndard d'aquest efecte. Per tal de realitzar la correcció sobre la desviació estàndard s'hauria de realitzar bootstrap.

2.7.3. Bootstrap

Un altre mètode que s'utilitza per tal de solucionar l'error de mesura és el bootstrap. Com ja s'ha comentat a l'apartat anterior, el mètode de regression calibration també fa servir un pas final de bootstrap per calcular els errors estàndards. Malgrat el bootstrap s'utilitza moltes vegades per obtenir estimacions d'errors estàndards, també es poden utilitzar les distribucions obtingudes amb el bootstrap per obtenir, per exemple, el valor mitjà d'un coeficient de regressió.

Com ja s'ha comentat, el bootstrap és un mètode àmpliament utilitzat per l'estimació de la variabilitat en mètodes estadístics complexos. La idea en sí és relativament senzilla. Es tracta de simular diferents bases de dades, anomenades "bootstrap data sets", basades en remostratge amb reposició de les dades originals, de manera que els bootstrap datasets tenen la mateixa distribució de probabilitat que les dades originals. Això ens permet realitzar qualsevol tipus d'anàlisi sobre els data sets simulats igual que si ho féssim sobre l'original. És per això que qualsevol distribució empírica de l'estimador o test estadístic realitzat sobre els data sets obtinguts mitjançant el bootstrap és vàlid per estimar la distribució del estadístic actual.

Per exemple, suposem que $\bar{\theta}$ és obtingut fent servir algun estimador de les dades originals, i que $\bar{\theta}^{(m)}$ s'obté d'aplicar el mateix estimador als bootstrap data set m , $m=1, \dots, M$, sent M el número de bootstraps generats, i $\bar{\theta}$ la mitjana de $\bar{\theta}^{(1)}, \dots, \bar{\theta}^{(M)}$. Aleshores, la matriu de variàncies-covariàncies de $\bar{\theta}$ es pot estimar a partir de:

$$\text{var}(\bar{\theta}) = (M - 1)^{-1} \sum_{m=1}^M (\bar{\theta}^{(m)} - \bar{\theta})(\bar{\theta}^{(m)} - \bar{\theta})^t$$

Tot i aquesta simplicitat, depenent del tipus de dades de les que es disposi i del tipus d'anàlisi a realitzar, el càlcul es pot tornar molt més complex. Hi ha diversos mètodes de càlcul de la distribució de probabilitat de les dades, i no sempre és obvi saber quin és el més adequat.

En general, el que es vol estimar quan es fa servir el bootstrap és un efecte o mitjana $\bar{\theta}$, així com una mesura de la precisió d'aquesta, ja sigui la desviació estàndard o el biaix. Més específicament, es vol conèixer la distribució del nostre estimador $\bar{\theta}$ al voltant del valor real θ . Aquest mètode es basa en la idea que la variabilitat de $\bar{\theta}$ al voltant de θ està determinada per

la variància de $\bar{\theta}^*$, que és l'estimació de la mitjana a partir dels bootstraps, al voltant de $\bar{\theta}$. Aquesta variabilitat inclou la variabilitat aleatòria (mesurada habitualment per la desviació estàndard) i la variació sistemàtica (biaix).

El bootstrap separa l'estimació en dos tipus: puntual, com es pot considerar l'estimació del biaix i l'error estàndard; i estimacions per intervals, que corresponen als límits inferiors i superiors. La desviació ja l'hem especificat anteriorment. El biaix es calcula de la forma:

$$\text{Biaix} = \bar{\theta}^* - \bar{\theta}$$

Pel que fa a l'estimació per interval s'ha de tenir en compte que és molt més informativa que l'estimació puntual, i que en molts casos, s'han proposat de fer servir mètodes de bootstrap corregits per la funció de probabilitat. D'estimació per interval n'hi ha de diversos tipus, (Basic bootstrap, Studentized bootstrap, normal aproximación, percentile bootstrap, i ajusted percentile bootstrap), tot i que totes elles es basen en la distribució dels percentils de les mostres del bootstrap. La mostra α -percentil del bootstrap ve donada per $(B+1)\alpha$, on l'ordre és aquell tal que $\bar{\theta}_1^* \leq \bar{\theta}_2^* \leq \dots \leq \bar{\theta}_B^*$. La diferència entre els diferents mètodes de bootstrap es basa en la tria de l'estadístic, o en la manera de convertir els percentils en intervals de confiança.

Aquesta tècnica també és emprada en altres contextos. En aquells casos en que la distribució teòrica és molt complicada, el bootstrap permet avaluar les propietats de la distribució i extreure'n les mostres i els paràmetres d'interès derivats d'aquesta distribució. També s'utilitza en els casos en que no es té suficient mostra com per poder realitzar inferència, o fins i tot, per conèixer la grandària mostral necessària quan es disposa d'una petita mostra, insuficient per dur a terme aquest objectiu.

En el capítol de metodologia explicarem específicament com apliquem el bootstrap en el context del nostre problema.

Capítol 3

Justificació de l'estudi

3.1. Origen de l'Estudi

Aquest projecte final de carrera es basa en l'estudi INMA (INfancia y Medio Ambiente), un estudi prospectiu de cohorts (Ribera d'Ebre, Menorca, Granada, Valencia, Sabadell, Asturias i Gipuzkoa). Es realitza el seguiment de 3500 nounats, des de la seva gestació fins a l'adolescència a diverses àrees geogràfiques, i es pretén estudiar el motiu d'anomalies en el nadó al néixer i el seu desenvolupament i la seva possible relació amb l'exposició a certs contaminants. D'aquesta manera, es planteja la possibilitat de descriure el grau d'exposició individual a diversos compostos orgànics i metalls durant la gestació i la primera infància; avaluar els efectes protectors i negatius de la dieta i el impacte de l'exposició a diferents contaminants sobre el creixement i el del desenvolupament del nen. Per desenvolupar aquest projecte ens centrarem en aquest últim punt, en concret, en l'exposició al NO₂ en les dones embarassades de la cohort de Sabadell i els seus efectes sobre el pes del nadó al naixement.

Les primeres cohorts en començar l'estudi varen ser Ribera d'Ebre i Menorca (1997), mentre que la cohort de Sabadell, estudiada en aquest projecte final de carrera, va iniciar el període de reclutament al 2004 (fins al 2006). Durant aquest període es van seleccionar dones majors de 16 anys, amb embaràs únic, sense malalties cròniques prèvies a l'embaràs (hipertensió crònica, arteriosclerosi, insuficiència renal crònica, vasculitis, càncer, alcoholisme, drogoaddicció), detectades en el programa d'atenció a gestants del Centre d'Atenció Primària especialitzada de Sabadell (CAP II Sant Felix) i que atorguessin el consentiment informat. Aquest centre tenia per norma fer la visita entre la setmana 4 i 6 de gestació, a més de ser l'únic centre d'atenció primària especialitzada de Sabadell.

A partir de la inclusió de les dones a l'estudi, es preveuen visites a les setmanes 20 i 32 de gestació, en el naixement, i quan el fill té 1 any, 4 anys, 8 anys i fins als 12 anys. Durant les diferents visites es recull informació referent a exposició a contaminants ambientals (Compostos orgànics volàtils (VOCs), partícules de matèria de diàmetre inferior a 10 µm (Pm₁₀) i Gasos (NO₂, CO, CO₂, entre d'altres), mesurats tots ells a diferents punts de la ciutat de Sabadell, així com metalls pesants (per exemple: Arsènic, Plom) i organoclorats (OC's, bifenils

policlorats (PCB) , hexaclorobenzè (HCB)) mesurats en sang, orina, i cordó umbilical, creixement intrauterí (Ecografies), ocupació de la mare (Qüestionari), i dieta (Qüestionari).

Els objectius generals de l'estudi INMA són:

1. Compartir metodologies i coneixements entre diversos grups espanyols que estudien els efectes del medi ambient a la salut infantil.
2. Descriure el nivell de contaminació i exposició dels nens durant la gestació i la primera infància
3. Avaluar el paper dels contaminants ambientals majoritaris i els factors protectors de la dieta en el creixement fetal i el desenvolupament neuro-endocrí-inmunitari.
4. Proveir informació útil per confeccionar programes indicadors de salut ambiental i d'avaluació de riscos.

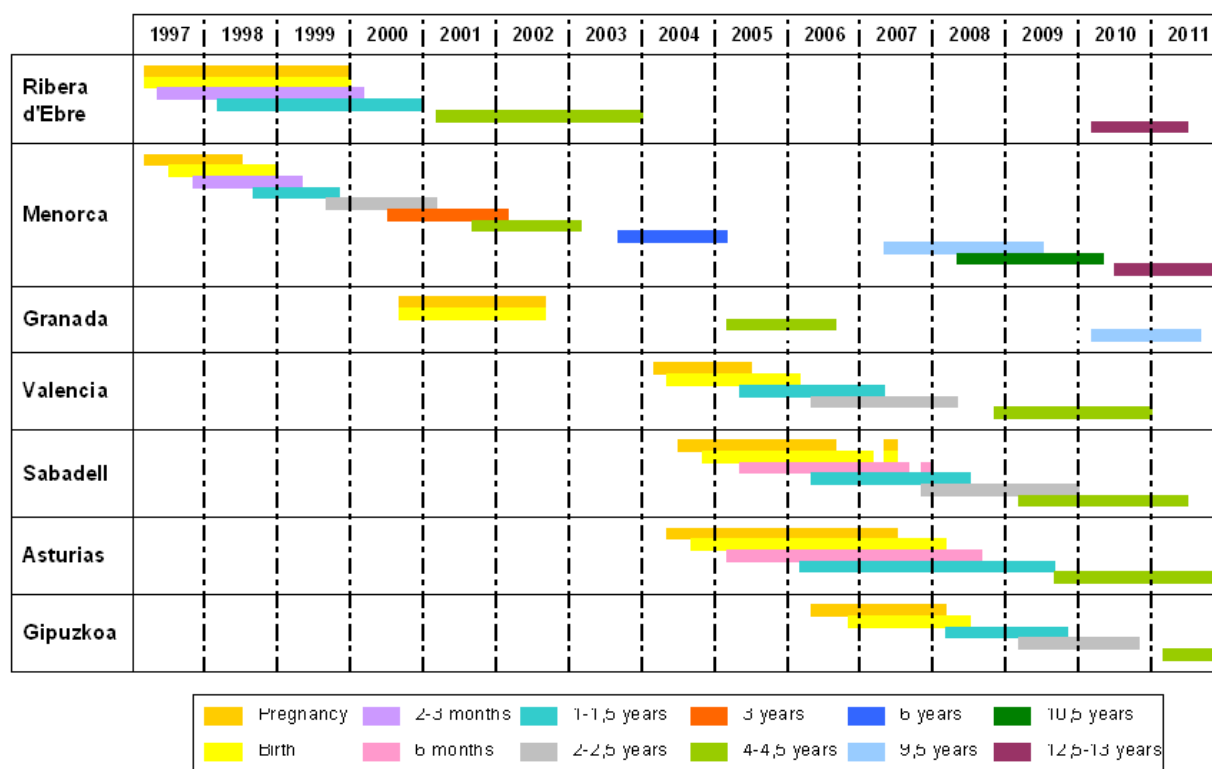
3.2. *Desenvolupament de l'Estudi*

Tot i que aquest estudi és multicèntric, cada cohort participant va tenir un període de reclutament diferent, que variava de 1 a 3 anys, i en diferents períodes (iniciant-se a Ribera d'Ebre al 1997, i l'últim, Gipuzkoa, al 2006). També va variar el tipus de reclutament, a les cohorts de Ribera d'Ebre i Granada es van reclutar durant l'admissió a l'hospital en el moment del part, mentre que a la resta de cohorts el reclutament es va produir prèviament al moment del part. A la taula 1 es mostra l'evolució de la població de les diferents cohorts del reclutament, al naixement del nadó. A Ribera d'Ebre, es van reclutar 102 dones en el moment del part. La cohort on van ser reclutades més dones va ser Valencia (855).

Àrees d'estudi	Anys d' inclusió	Dones embarassades	Nounats
Ribera d'Ebre	1997-1999	102	102
Menorca	1997-1998	482	482
Granada	2000-2002	668	668
Valencia	2004-2005	855	787
Sabadell	2004-2007	787	748
Asturias	2004-2007	590	485
Gipuzkoa	2006-2008	639	613

Taula 1. Reclutament per cohort

A la següent taula podem observar l'evolució esperada de les diferents cohorts per any de seguiment.



Taula 2. Seguiment per cohort

3.2.1. Resultats obtinguts

Com a conseqüència de les anàlisis realitzats durant l'estudi s'han generat, fins ara, més d'una vintena de publicacions sobre els diferents temes d'estudi. A continuació comentaré algunes de les publicacions realitzades relacionades amb els objectius del projecte final de carrera.

Una de les primeres publicacions va ser la realitzada per Sunyer et al. En aquesta es buscava la possible relació entre el NO₂ mesurat a l'interior de 1611 llars de Barcelona, Menorca i Ashford (Anglaterra) i la presència d'infeccions respiratòries (clínicament diagnosticades) durant el primer any de vida. En aquest cas, les dades no provenien només de l'estudi INMA, sinó que es van extreure de l'estudi AMICS (Asthma Multicenter Infants Cohorts Study) en el qual participava la cohort de Menorca. En les diferents anàlisis realitzades no es va trobar associació estadística entre l'exposició al NO₂ a l'interior de les llars i la presència d'infeccions respiratòries.

Jordi Sunyer, Carme Puig, Maties Torrent, Oscar Garcia-Algar, Ignasi Calicó, Laura Muñoz-Ortiz, Maria Barnes, Paul Cullinan on behalf of the AMICS study..Nitrogen dioxide is not associated with respiratory infection during the first year of life.International Journal of Epidemiology.2004.PMID: 15075156

Un altre article publicat al 2007 per Ana Esplugues et al. es basava en el protocol a seguir per les diferents cohorts a l'hora de mesurar els diferents tipus de contaminants, ja fossin a l'aire lliure (NO₂, VOC's,...), a l'interior de les llars, o en orina. També es feia referència als mètodes de recollida de la informació relacionada amb els contaminants mesurats (sistemes d'informació geogràfica o bé, mitjançant qüestionaris).

A. Esplugues, R. Fernández-Patier, I. Aguilera, C. Iñíguez, S. García Dos Santos, A. Aguirre Alfaro, M. Lacasaña, M. Estarlich, J. O. Grimalt, M. Fernández, M. Rebagliato, M. Salah, A. Tardón, M. Torrent, M.D. Martínez, N. Ribás-Fitó, J. Sunyer, F. Ballester. Exposición a contaminantes atmosféricos durante el embarazo y desarrollo prenatal y neonatal: protocolo de investigación en el proyecto INMA (Infancia y Medio Ambiente).

Per últim, destacar dos dels articles en els quals es basa aquest projecte, i que van ser realitzats per Inmaculada Aguilera et al. En el primer d'ells s'explica el mètode utilitzat per trobar models de predicció de diferents contaminants ambientals (NO₂, BTEX entre d'altres) a la ciutat de Sabadell. En aquest article s'explica com es van obtenir aquests models a partir de la mesura presa en 57 punts de la ciutat de Sabadell de NO₂, BTEX i altres contaminants, i de la informació geogràfica i de trànsit recollida a través del software informàtic ArcGis.

Aguilera I, Sunyer J, Fernández-Patier R, Hoek G, Aguirre-Alfaro A, Meliefste K, Bomboi-Mingarro MT, Nieuwenhuijsen MJ, Herce-Garraleta D, Brunekreef B. Estimation of outdoor NOx, NO2 and BTEX exposure in a cohort of pregnant women using Land Use Regression modeling. Environ Sci Technol 2008; 42 (3): 815-21.

En el segon article al que faig referència de Aguilera et al, es presenten els resultats obtinguts d'estudiar l'efecte de l'exposició al NO₂, i a altres contaminants (predits per cadascuna de les dones de la cohort de Sabadell mitjançant el model LUR explicat a l'article anterior) durant els diferents trimestres de l'embaràs i el pes del nadó al néixer. En cap cas es va trobar associació estadística entre els diferents nivells d'exposició dels diferents contaminants estudiats i el pes del nadó. L'únic efecte significatiu que es va trobar, va ser un retard en el creixement intrauterí del nadó al segon trimestre d'embaràs, relacionat amb l'exposició al BTEX durant aquest període.

Association between GIS-based exposure to urban air pollution during pregnancy and birth weight in the INMA Sabadell Cohort. Aguilera I, Guxens M, Garcia-Esteban R, Corbella T, Nieuwenhuijsen MJ, Foradada CM, Sunyer J. Environ Health Perspect. 2009 Aug;117(8):1322-7. Epub 2009 Apr 13.

Com a conseqüència del gran volum d'informació que genera l'estudi INMA, i amb la intenció de facilitar l'accés a tota la informació extreta d'aquest estudi, durant l'any 2009 es va crear la pàgina web de l'estudi (www.infanciaymedioambiente.org) . En aquesta web es pot trobar una descripció de l'estudi, així com les diferents publicacions que es van produir.

3.2.2. Vincle entre contaminació atmosfèrica i Pes al néixer

La necessitat d'avaluar els efectes de la contaminació sobre les dones embarassades i la salut del nadó al néixer prové de l'evidència científica trobada als darrers anys en diversos estudis. Se sap que els primers anys de vida són claus pel desenvolupament dels pulmons i del sistema immunitari, i per tant, l'exposició a la contaminació atmosfèrica durant aquest període pot ser més important. S'ha de tenir en compte que els nens respiren més aire que els adults respecte a la seva massa corporal i, a més, estan més temps realitzant activitats a l'aire lliure, pel que l'exposició en edats infantils és major (Olsen,2000; Schwartz, 2004) que en edats adultes.

La majoria d'estudis s'han basat en l'estudi dels efectes perinatals més rellevants, com són el creixement fetal i els naixements prematurs. S'utilitza el terme creixement intrauterí retardat per indicar un desenvolupament del fetus per sota del seu potencial gnòmic de desenvolupament, degut a un procés patològic durant l'etapa prenatal. Però degut a la falta de definició clínica d'aquest fenomen, diversos indicadors s'han fet servir, com són el pes del nadó al néixer com a variable contínua, baix pes al néixer (<2500 g), molt baix pes al néixer (<1500 g) o petit per l'edat gestacional (PEG), que es defineix com el pes inferior al percentil 10 per una edat gestacional determinada. Aquesta última mesura, al ser totalment depenent de la població en estudi, s'ha considerat que no era la més adequada. Per contra, s'ha considerat la desviació del potencial de creixement al néixer i no el pes al néixer en valor absolut com la millor aproximació. Un altre problema que es planteja al voler avaluar el pes al néixer respecte els nivells d'exposició als quals ha estat exposada la mare és el fet de la necessitat de tenir coneixement d'altres covariables que poden intervenir en el desenvolupament normal del nadó, i que potencien la necessitat de realitzar estudis de cohorts.

Un altre dels punts a tenir en compte, donat que l'embaràs és un període curt i determinat, és el d'avaluar quin és el període de major susceptibilitat, si existeix, a l'exposició ambiental. Però aquest punt és complex degut a que diferents contaminants poden tenir efectes en diferents períodes, o bé, que els contaminants mesurats poden ser únicament indicadors dels autèntics contaminants que tenen l'efecte, o, un cop més, a la variabilitat dels contaminants tant espacial com temporal.

3.2.3. Importància dels estudis epidemiològics

Un dels condicionants més importants en epidemiologia i tret diferencial respecte altres àmbits és el fet que els estudis que es realitzen són observacionals. El fet de dissenyar un estudi no inclou la possibilitat de realitzar un disseny d'experiments. S'ha de tenir en compte que habitualment s'estudia l'efecte de certs components nocius per la salut sobre els propis humans, i per tant, no seria gaire ètic obligar a un determinat grup de persones a estar exposades a una component nociva per la seva salut, a diferència del que a vegades es fa amb animals. Aquest fet també comporta problemes de caracterització de l'exposició. En moltes ocasions, l'epidemiologia és l'eina emprada en casos d'extrema urgència. Fa pocs anys, a la ciutat de Barcelona es va registrar un increment d'episodis d'asma en un breu període. Gràcies a l'estudi epidemiològic que es va realitzar es va descobrir que aquests brots eren deguts a la descàrrega, al port de Barcelona, de gra de soja.

Es per això pel que l'epidemiologia ambiental pren una gran importància en l'àmbit clínic. No es podrien vincular algunes de les causes de malalties respiratòries o cardiovasculars sense unes bones mesures de contaminació.

Capítol 4

Objectiu d'aquest projecte

L'objectiu final del projecte és avaluar l'efecte de la contaminació ambiental en el pes del nadó al néixer. Per tal de desenvolupar el projecte disposarem de dues bases de dades: la primera, amb informació de contaminació ambiental a diversos punts de la ciutat de Sabadell; la segona amb informació de les dones de l'estudi, de les quals també disposem de la seva adreça.

En aquest projecte final de carrera es pretén trobar un mètode estadístic per tal de millorar l'estimació de l'efecte de l'exposició durant el període d'embaràs al diòxid de nitrogen (NO_2) i el pes del nadó al néixer. Per fer-ho, s'inclourà la incertesa de les prediccions de l' NO_2 per cada dona en el model. D'aquesta manera s'espera reduir el possible biaix i l'atenuació de l'efecte que en estudis previs es dona a l'avaluar l'efecte de l'exposició ambiental sobre algunes malalties, i corregir els intervals de confiança per tenir en compte aquesta font addicional de variació. Aquests anàlisis van ser realitzats prèviament per Aguilera et al. (comentats a la secció 2.2.1. Resultats obtinguts) pel que el desenvolupament del projecte serà similar, tot i que s'incorporaran diferents tècniques estadístiques per tal de, amb les mateixes dades, millorar el procés tant de predicció com d'estimació de l'efecte.

Els objectius es centraren en tres punts:

- Trobar un bon model per l' NO_2 mesurat a diferents punts de la ciutat de Sabadell i fer una bona predicció per a les dones participants de l'estudi i avaluar les prediccions.
- Estimar l'efecte de l'exposició a l' NO_2 en el pes del nadó al néixer.
- Donat que els valors de NO_2 són predits i no observats, implementar un mètode per incorporar la incertesa de la predicció del NO_2 en l'anàlisi respecte el pes al néixer i comparar-ho amb els models que no incorporen aquesta incertesa.

Aquest últim punt dels objectius ve motivat pel fet que en un anàlisi futur de les dades de l'estudi ARIBA, explicat a la secció 2.2, es vol realitzar un anàlisi similar a aquest incorporant

aquests conceptes, en el qual es voldrà estudiar la relació entre l'exposició a certs contaminants ambientals i la presència de determinades anomalies dels nounats al néixer.

Capítol 5

Recollida de dades de l'estudi

Introducció

Tot i que l'estudi INMA és un estudi multicèntric, per desenvolupar aquest projecte únicament es va fer servir la informació referent a la cohort de Sabadell. Al inici del mateix no es disposava de totes les dades necessàries per poder realitzar l'anàlisi conjunt. En cas que s'haguessin fet servir la resta de cohorts els models de predicció del diòxid de nitrogen i d'ajust del pes del nadó al néixer haguessin estat molt més complexos. Així doncs, el nostre estudi parteix d'una N inicial de 657 dones.

Fonts d'informació

En aquest estudi es va recollir informació de 4 tipus:

- Variables relacionades amb hàbits de la llar i personals al total de la població en estudi de la ciutat de Sabadell (entre d'altres, el temps que passa fent activitats fora i dins de casa, hàbit tabàquic, temps ventila la llar o si té cuina de gas). De totes les variables recollides en aquest estudi, únicament farem servir les variables estat civil de la mare, ètnia, classe social de la mare, nivell d'estudis, tabaquisme, tipus de cuina, edat de la mare, pes, alçada i IMC (Índex de massa corporal, calculat com el pes en quilograms partit per l'alçada en metres al quadrat) de la mare previ a l'embaràs i número d'embarassos previs.
- Variables referents al nadó com són el pes o l'alçada en el moment de néixer i durant el seu creixement (Bayley's Test). En aquest projecte, com a variable resposta dels models, farem servir el pes del nadó al néixer. També es faran servir les variables setmanes de gestació, el sexe del nadó i l'època de gestació en l'ajust del model del pes al néixer.
- Variables de mesura de la contaminació de NO₂, aquestes dades varen ser recollides a 57 punts de mostreig de la ciutat de Sabadell, i a la llar (dins i fora) d'una petita mostra de dones de l'estudi (57). També es va recollir informació d'una estació central de trànsit. En el nostre cas únicament farem servir les

mesures preses a 57 punts de la ciutat de Sabadell, així com les dades provinents de l'estació de trànsit.

- Variables geogràfiques, variables relacionades amb el trànsit (per exemple: carrer molt transitat, o la distància a benzineres) i tipus de terreny (rural, urbà, industrial), que varen ser recollides en els punts on es va mesurar l'NO₂ i a les llars de les dones de l'estudi (611, 46 dones no vivien a Sabadell). De totes les variables que es varen recollir referents a aquest punt únicament utilitzarem les coordenades X i Y (UTM), que és un sistema de coordenades basat en projeccions cartogràfiques Mercator (projecció que conserva els angles i les formes). Aquestes projeccions es calculen partir del meridià, i no a partir de l'equador a diferència del sistema de coordenades geogràfiques; l'alçada del punt respecte al nivell de l'aigua pel que fa a variables geogràfiques, i el tipus de carrer (molt, mitjà o poc transitat), el tipus de terreny (urbà, industrial o rural) i altres variables de trànsit.

5.1. Dades de NO₂

La informació d'exposició al NO₂ es va obtenir en funció del lloc de mesura (en 57 punts de la ciutat de Sabadell, i en 1 estació de trànsit). Per tal d'obtenir aquestes mesures es van fer servir 2 tipus diferents de captadors, Radiello i Ogawa. Cadascun d'aquests captadors es va fer servir en 2 de les 4 campanyes de recollida de dades realitzades. Per tal d'evitar diferències entre els captadors, i degut a que semblava que existia una diferència sistemàtica entre les mesures preses pels dos captadors (en mateixos llocs, en el mateix període) a la mesura del captador d'Ogawa se li va aplicar una correcció a partir de la regressió lineal de la mesura de Radiello = $8.034812 + (0.9805348 * \text{Ogawa})$ (R^2 de la regressió era de 0.85) (Aguilera et al. 2008). A la taula 3 podem observar el període en el que es van recollir les diferents mostres, la mesura recollida i el punt de recollida.

Abril 2005	Juny 2005	Octubre 2005	Març 2006	Número de mostres
Mitjana de 7 dies	Mitjana de 7 dies	Mitjana de 7 dies	Mitjana de 7 dies	57 punts
Dades diàries de Trànsit (1 juny 2001 fins 31 març del 2007)				1 Estació de trànsit

Taula 3. Diferents fonts d'informació del NO₂ i període de recollida

5.1.1. Mostreig a 57 punts de Sabadell

Les primeres mesures preses de NO₂ varen ser la dels 57 punts de Sabadell. Aquests punts es van distribuir de forma que es pogués recollir la possible variabilitat espacial del NO₂

a Sabadell. A la figura 3 es pot observar la distribució dels punts de mostreig on es van realitzar la recollida de dades, així com la situació de les vivendes de les dones participants a l'estudi. Durant aquest procés de recollida es va obtenir 1 mesura durant 4 períodes de l'any (19-26 Abril 2005, 31-7 Juny 2005, 18-25 Octubre 2005, i 14-21 Març 2006). En cadascun d'aquests períodes es va recollir la mitjana de NO₂ dels 7 dies mesurats.

L'objectiu de mesurar la contaminació de NO₂ durant 7 dies en diferents períodes és per poder fer una predicció de la mitjana diària d'exposició al NO₂ per qualsevol punt de la ciutat. Aquesta predicció estarà basada en un model de regressió lineal que utilitza com a variables explicatives variables d'usos del sòl referents a la localització del punt (p.ex. densitat de trànsit en aquell punt o l'amplada del carrer). Com que la informació d'aquestes variables està disponible per a qualsevol punt de la ciutat, el model resultant es pot fer servir per predir l'NO₂ en qualsevol punt. En epidemiologia, a aquesta estratègia se l'ha batejat com *Land Use Regression Model* (LUR model), com ja hem explicat a l'apartat 1.4.4.

Resumint, en aquest primer procés de recollida de mesures d'exposició al NO₂ obtenim el promig d'exposició al NO₂ dels llocs de mostreig pel model (57 llocs) (a partir d'ara anomenarem NO₂ LUR)

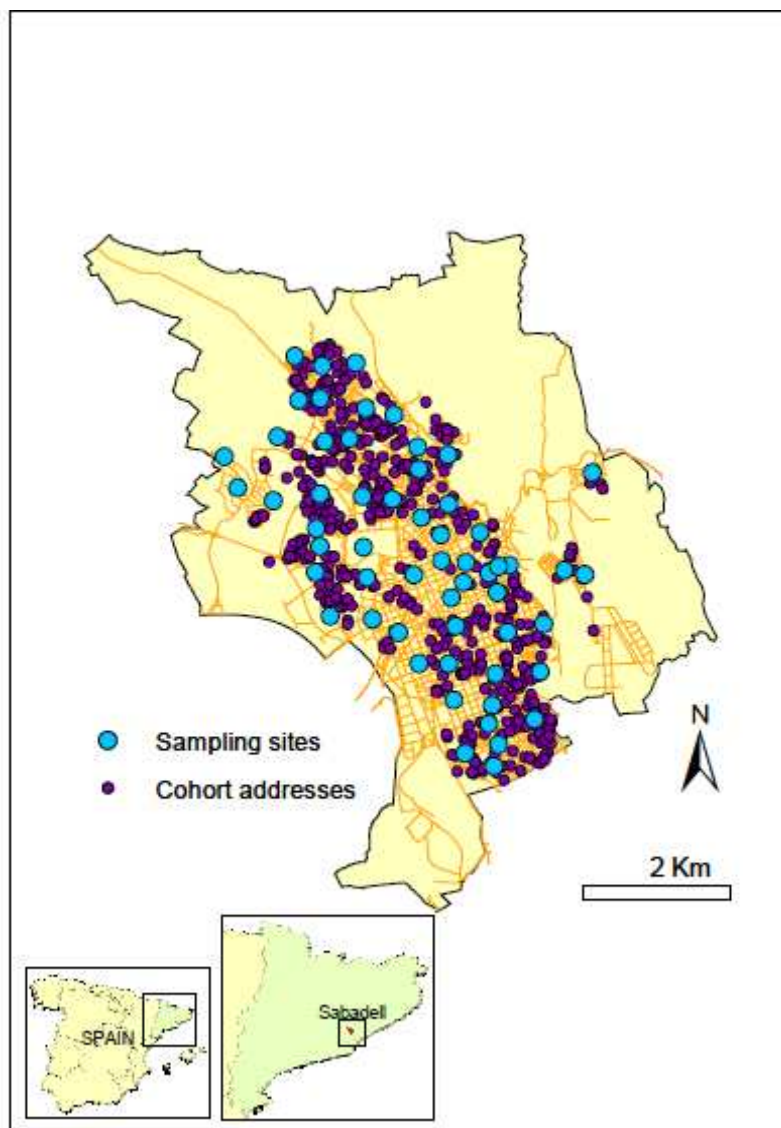


Figura 3. Punts mostrejats de la ciutat de Sabadell

5.1.2. Dades de l'estació de trànsit

La segona font d'informació referent al NO_2 eren les mesures recollides a una estació de trànsit de Sabadell. Aquesta estació proporciona dades diàries de la mitjana de NO_2 des de l'1 Juny del 2001 fins al Març del 2007.

5.2. Dades de Trànsit i geogràfiques de Sabadell

Existeixen bases de dades georeferenciades que donen informació sobre variables associades a una determinada coordenada en un mapa. Ajuntaments i altres institucions disposen d'aquests tipus de bases de dades, de on es poden extreure, per exemple, els usos del sòl (urbà, rural o industrial), dades referents al volum de trànsit, distància a benzineres, pàrkings, o dades de l'alçada geogràfica. El maneig d'aquestes bases de dades es va fer amb

el software informàtic (ArcGIS) que, a partir de les coordenades que se li indiquen, et retorna informació referent a les diferents variables.

5.3. Dades de les dones de l'estudi i del nadó al néixer

Al total de les dones (657), se'ls hi passen durant diversos moments de l'embaràs qüestionaris referents a les condicions de la llar, hàbits de neteja, hàbit fumador, dieta entre d'altres. A l'annex 1 es pot trobar un dels qüestionaris fet servir durant l'estudi per tal de recollir la informació.

Durant el nostre estudi únicament farem servir les variables obtingudes del qüestionari de la setmana 12 d'embaràs referents a dades generals de la mare, així com un qüestionari ambiental, referent a les condicions de la llar i hàbits de neteja.

Pel que fa a la informació del nadó i al procés de gestació, durant l'embaràs es recull informació referent a l'embaràs, ecografies, historial de malalties, i tot allò que tingui a veure amb medicacions durant aquest període. D'aquesta manera es pot determinar l'evolució del nadó durant el període d'embaràs, així com situar en el temps, en cas que es donin, l'existència de complicacions durant l'embaràs. A més, en el moment del part, es va recollir tota la informació perinatal, és a dir, el pes, l'alçada, el perímetre cranial entre d'altres.

Capítol 6

Metodologia

Per tal de desenvolupar els objectius de l'estudi farem servir dues metodologies diferents: model lineal, i model lineal amb bootstrap. En cadascun d'aquests dos mètodes d'anàlisi haurem de realitzar 3 passes ben diferenciades.

La primera part consistirà en modelar les 57 mesures de diòxid de nitrogen preses a la ciutat de Sabadell en funció de les variables geogràfiques i de trànsit. Dintre de les variables geogràfiques farem servir les coordenades UTM x (longitud) i UTM y (latitud) amb la idea que poden capturar un patró especial de l'NO₂; també s'utilitzarà l'altitud del punt de mostreig respecte al nivell del mar. Igualment, es faran servir les variables de densitat de població, i densitat d'edificis tenint en compte diferents radis de distància des del punt de mostreig, així com el tipus de terreny que l'envolta (terreny urbà, industrial o rural). Dintre de les variables de trànsit s'inclouran variables que contenen informació referent al tipus de carrer en el que es troba el punt de mostreig, definit en funció del volum de trànsit que s'hi registra, alt, mitjà o baix. També els km de vies amb alt, mitjà o baix volum de trànsit que hi ha en un radi de 50 i 500 m des del punt de mostreig. Per últim, es consideraran variables referents a la distància a possibles fonts de NO₂, pàrkings, tallers o benzineres. En aquesta part de l'anàlisi aplicarem cross validation per tal de conèixer la qualitat del model de predicció. Farem servir el mètode 10-fold, descrit al capítol 2.5.

Un cop modelat l'NO₂, i com a segona part de l'anàlisi, amb el model derivat de la primera fase es realitzarà la predicció de l'NO₂ a l'adreça de residència de les 611 dones pertanyents a l'estudi, de les quals es disposa de la informació geogràfica i de trànsit de la seva residència.

Per finalitzar, es realitzarà l'ajust del pes al néixer del nadó respecte a variables referents a característiques de la mare (classe social, edat, educació, pes i alçada abans de l'embaràs, estat civil, ètnia, número de parts previs, època de l'embaràs), hàbits durant el període d'embaràs (hàbit tabàquic, el tipus de cuina), característiques del nadó al néixer (el sexe i l'edat gestacional), així com la predicció de l'exposició a l'NO₂ durant el període d'embaràs. És en aquest punt on es vol tenir en compte que l'NO₂ pren un valor predit i no el valor real.

A la figura 4 podem veure el procés que es desenvolupa per tal de poder estimar l'efecte de l'exposició al diòxid de Nitrogen durant el període d'embaràs sobre el pes del nadó al néixer.

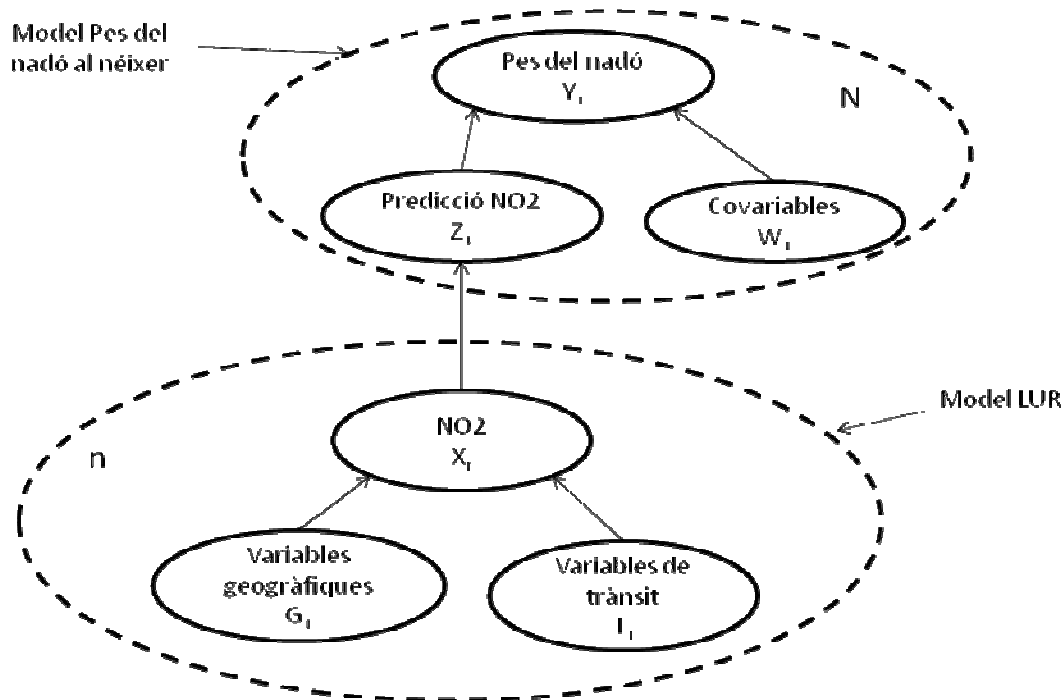


Figura 4. Diagrama de l'ajust dels models de predicció de l'NO₂, i del model pes al néixer incorporant la incertesa de la predicció

Els model emprat per predir l'exposició al NO₂ serà el model lineal. Tot i que existeix una variabilitat temporal en les mesures ambientals, ja que es van prendre quatre mesures diferents a cada lloc en diferents moments en el temps, es farà servir com a variable resposta la mitjana de les quatre mesures, ponderada per la tendència temporal observada en l'estació fixa que disposa de dades diàries. Pel que fa al model del pes al néixer, igual que en el cas anterior, s'ajustarà un model lineal multivariant. Per tal de tenir en compte la incertesa de les prediccions, utilitzarem el mètode freqüentista del bootstrap. Aquest procés es descriu al capítol 6.2.

Al final de l'anàlisi compararem els resultats obtinguts sense tenir en compte la incertesa deguda a l'error de mesura (model lineal), o sigui, ajustant els dos models lineals (model LUR i model Pes del nadó al néixer, Figura 4) de forma independent, i tenint en compte l'error de mesura mitjançant el bootstrap, de manera que la incertesa del model LUR es transmet al model del pes del nadó al néixer.

6.1. Model lineal

La regressió lineal és una de les eines estadístiques més utilitzades a l'hora d'analitzar dades. Molts estudis científics es refereixen a la "relació" com el que succeeix entre dues o més quantitats observades. Com una quantitat y , varia en funció d'una altra quantitat o vector x .

En els models lineals la clau de la modelització és que la relació entre les variables x i y sigui lineal (amb possibilitat de transformacions), amb residus independents i aproximadament normals, amb mitjana zero i variància constant.

Un model lineal normal es pot escriure de la següent forma matricial,

$$y = X\beta + \varepsilon,$$

on y és el vector $n \times 1$ d'observacions de la resposta, X és la matriu de coeficients coneguts $n \times p$, β és el vector de paràmetres $p \times 1$ i ε és el vector d'errors $n \times 1$. Els elements de ε s'assumeixen que tenen valor esperat i correlació entre ells igual a 0, variància constant σ^2 , i es distribueixen segons una normal. És a dir, que $\varepsilon \sim N(0, I\sigma^2)$. El model únicament diu que la distribució de y condicionada als paràmetres (β, σ^2) és la distribució normal multivariant:

$$y | \beta, \sigma^2, X \sim N(X\beta, I\sigma^2),$$

on I es la matriu identitat $n \times n$.

Partint de la distribució multivariant normal, la funció de versemblança s'escriurà de la següent manera:

$$l(\beta, \sigma^2 | y) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2} \right\}$$

on la forma quadràtica $(y - X\beta)'(y - X\beta)$ es pot escriure com:

$$(y - X\beta)'(y - X\beta) = \beta'X'X\beta - \beta'X'y - y'\beta X + y'y$$

L'estimació de paràmetres més utilitzat pel model lineal és el de mínims quadrats:

$$\hat{\beta}_{LS} = (X'X)^{-1}X'Y$$

I donat que estem treballant amb un model lineal normal, l'estimador de mínims quadrats coincideix amb el de màxima versemblança. Aquest, és l'estimador de mínima variància dels estimadors sense biaix.

Aleshores, si el model és correcte la distribució que segueix aquest estimador és:

$$\tilde{\beta}_{MV} | \beta, \sigma^2 \sim N(\beta, \sigma^2 (X'X)^{-1})$$

La regió de confiança per β de $(1-\alpha)$ estarà format per les β 's que compleixen:

$$(\tilde{\beta}_{MV} - \beta)'(X'X)(\tilde{\beta}_{MV} - \beta) \leq p S_R^2 F(\alpha; p, n - p),$$

On p és el número de variables explicatives, S_R^2 la suma de quadrats residual, i F és la distribució F-Snedecor.

6.2. Bootstrap

En el nostre cas, el procés de remostratge ha de tenir en compte les dues bases de dades amb les que treballem per capturar la variabilitat dels dos models, tant a la del model per predir l'NO₂ com a la del model del pes al néixer en funció del NO₂ predit. L'algorisme que hem desenvolupat consisteix en tres passes. Crea un bootstrap de la base de dades del NO₂, i ajusta un model on l'NO₂ es la variable depenent i les variables geogràfiques i de trànsit les variables independents (model 1). Posteriorment, realitzar la predicció del NO₂ sobre el bootstrap de la base de dades del pes del nadó. Per últim, s'ajusta el model del pes al néixer (model 2) fent servir el bootstrap del pes amb la nova predicció del NO₂. A la figura 5 es pot observar gràficament aquest procés, iterat 200 vegades.

D'aquesta manera s'aconsegueix transmetre la incertesa del model 1 al model 2, donat que en cada bootstrap dataset el model 1 dóna coeficients una mica diferents, i per tant, també prediccions de l'NO₂ diferents, que llavors s'entren com a variables explicatives al model 2. Així doncs, la distribució de les 200 estimacions de l'efecte de l'NO₂ contindrà una part de la variació deguda a les variacions de la predicció de l'NO₂, és a dir, a la incertesa de la predicció de l'NO₂.

corresponent al pes, que contenia 300 individus, s'ha de distingir entre l'escenari 1, on simulàvem una relació significativa amb el pes, i l'escenari 2 on aquesta relació no existia. En el primer escenari es va repetir el procés anterior. Es va generar una variable `no2.unobserved`, que representa el valor real (no observat en l'estudi) de l' NO_2 per tots els participants de la següent manera: $\text{no2.unobserved} = 1.5 + 2.6 * x1.\text{weight} + \text{rnorm}(n=50, \text{mean}=0, \text{sd}=1.8)$. on `x1.weight` representa el valor de la variable `x1` en la població d'estudi. `x1.weight` es va simular amb la mateixa distribució que `x1`, ja que en ambdós casos representa la mateixa variable, la variable que es farà servir per predir l' NO_2 a la població. Després, assumint que l' NO_2 real influeix en el pes, es va crear la variable de pes `weight` en base a la següent equació, $\text{weight} = 1 + 2 * \text{no2.unobserved} + \text{rnorm}(n=300, \text{mean}=0, \text{sd}=1.5)$, de manera que ens asseguràvem que la relació entre la predicció generada a partir del model ajustat a la base de dades del NO_2 tindria un efecte significatiu sobre el pes al néixer. A l'escenari 2 vàrem realitzar el mateix procés, canviant el coeficient de `no2.unobserved` a 0.05 per tal que no fos significativa. En aquestes dades simulades es va realitzar l'anàlisi habitual i també l'anàlisi bootstrap descrit a la figura 5 (A l'annexa 2 es pot trobar la sintaxi utilitzada).

A la taula 4 podem observar els valor obtinguts al realitzar el model lineal i el model lineal amb bootstrap per a cadascun dels dos escenaris creats. Si ens fixem a l'escenari on l' NO_2 sí que tenia relació amb el pes s'observa com al realitzar el model lineal amb el bootstrap s'obté un efecte similar, però uns intervals de confiança molt més amplis. Això es deu a que al fer servir aquest mètode incorporem la incertesa del model de predicció a l'estimació de l'efecte del NO_2 sobre el pes. Per contra, si ens fixem als resultats obtinguts per l'escenari on el diòxid de nitrogen no tenia relació amb el pes s'observa que la diferència en l'amplitud d'intervals de confiança pels dos mètodes emprats són molt més semblants, tot i que també mostra us intervals lleugerament més amplis.

	NO_2 relacionat amb el pes	NO_2 sense efecte en el pes
	Coeficient (IC 95%)	Coeficient (IC 95%)
Regressió lineal	2,36 (2,17, 2,55)	0,13 (-0,02, 0,28)
Regressió lineal i Bootstrap	2,41 (1,770, 2,892)	0,16 (-0,01, 0,41)

Taula 4. Comparació de les estimacions obtingudes pel model lineal i el model lineal més bootstrap de les dades simulades en funció de si l' NO_2 tenia o no efecte en el pes

Per tal d'avaluar si els intervals de confiança tenen el *coverage* correcte, és a dir, si inclouen el valor real un 95% de les vegades, s'han simulat 200 bases de dades del NO_2 i del pes seguint les condicions de l'escenari 1 l'escenari 2 (descrits anteriorment). En cada una de les 200 bases de dades s'han realitzat les anàlisis ajustant el model 1 i el model 2 de forma independent, i també fent servir el procediment bootstrap descrit anteriorment, guardant els intervals de confiança de l'efecte estimat de l' NO_2 respecte el pes en cada base de dades i per

cada mètode i escenari. D'aquesta manera vàrem poder constatar si el *coverage* dels dos mètodes era el correcte, és a dir, si el 95% de les vegades el interval de confiança estimat contenia el valor real. Per l'escenari 1, al ajustar el model lineal es va obtenir que únicament en un 64% dels casos els intervals de confiança varen incloure el valor real, mentre que quan es feia servir el procediment bootstrap aquest percentatge s'incrementava fins al 94.5% (figura 6). Al realitzar el mateix procediment sobre l'escenari 2, es va obtenir que tots dos casos, model lineal i model lineal amb bootstrap el percentatge era igual, un 94% (figura 7). Confirmant el fet que el bootstrap, en cas que no existeixi efecte, no aporta res a l'anàlisi, i en cavi, si existeix efecte, es converteix en una eina necessària en cas d'error de mesura.

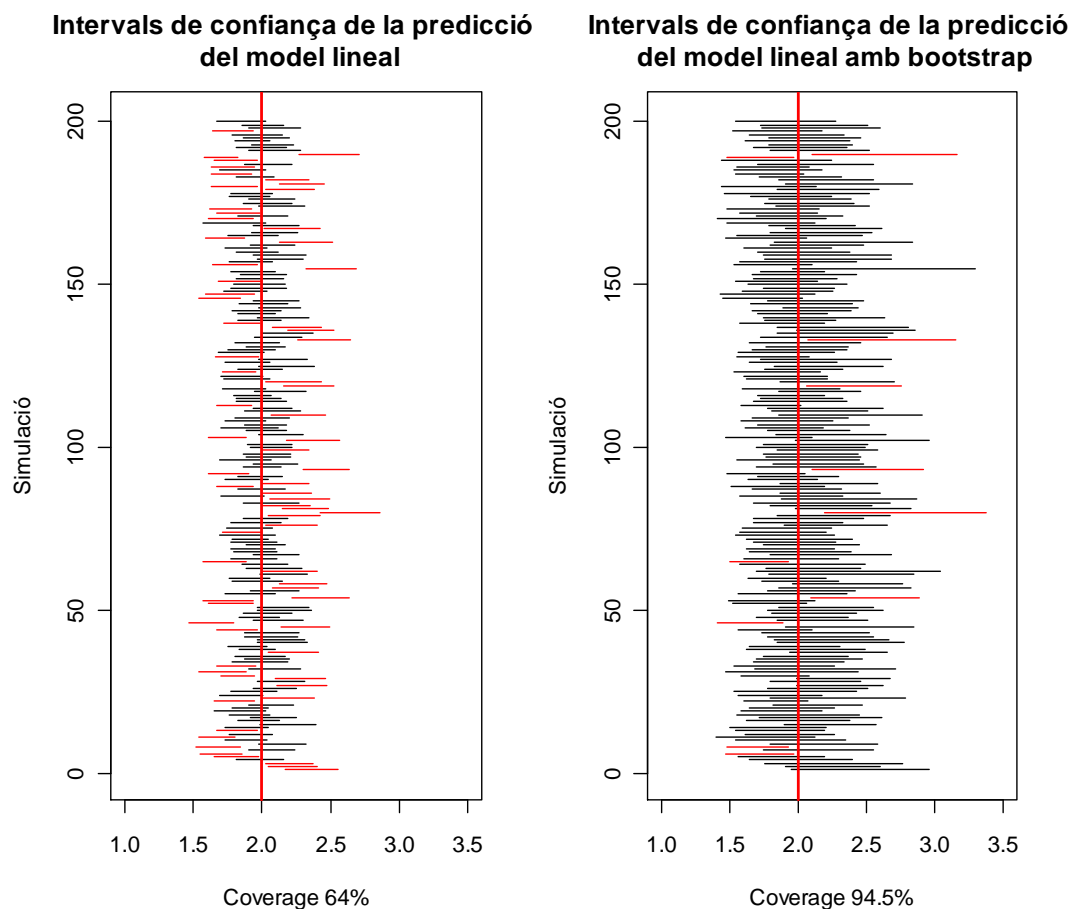


Figura 6. Gràfic dels intervals de confiança predits a l'escenari 1 pel model lineal amb el valor real de la predicció (dreta) i gràfic dels intervals de confiança predits pel model lineal amb bootstrap amb el valor real de la predicció. En vermell aquells intervals que no contenen l'efecte real.

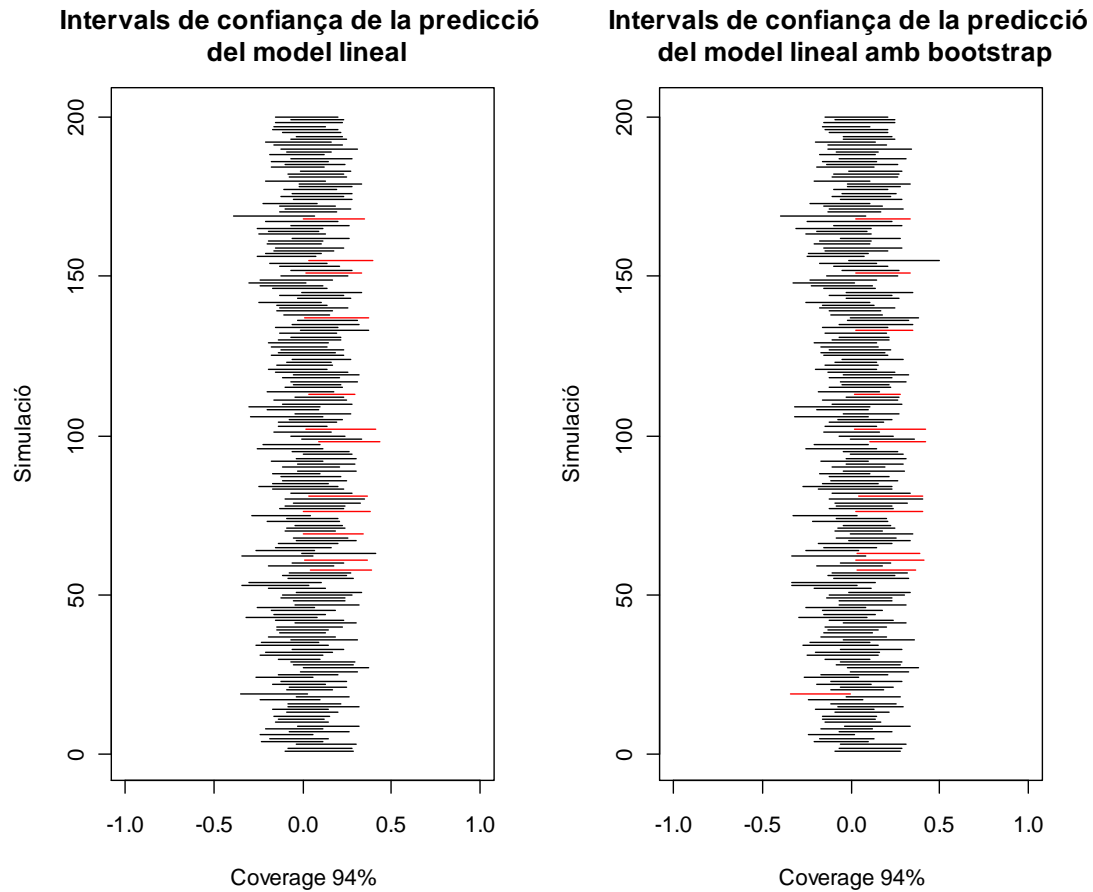


Figura 7. Gràfic dels intervals de confiança predits a l'escenari 2 pel model lineal amb el valor real de la predicció (dreta) i gràfic dels intervals de confiança predits pel model lineal amb bootstrap amb el valor real de la predicció. En vermell aquells intervals que no contenen l'efecte real.

Capítol 7

Descriptiva de les dades

7.1. Dades NO_2

7.1.1. Mesures de NO_2 preses a 57 punts de Sabadell

A les figures 8-11 observem la disposició espacial dels punts mostrejats i els valors mitjos de diòxid de nitrogen recollits en els diferents períodes de mostreig (taula 5). S'observa com els valors més alts en els diferents punts mostrejats es van donar al mesos d'abril i juny, i com eren als punts centrals de la ciutat de Sabadell on s'observa que a totes les campanyes es van acumular les mesures més elevades, tot i que sense ser del tot clara. Es veu una mena de pic en els punts centrals del gràfic (coincidents amb el centre de la ciutat) i un descens als extrems del valors de NO_2 . S'ha de tenir en compte que la ciutat de Sabadell té una forma allargada que capturen els punts de mostreig del gràfic. Així doncs, sembla que el patró espacial es manté entre les diferents campanyes registrades.

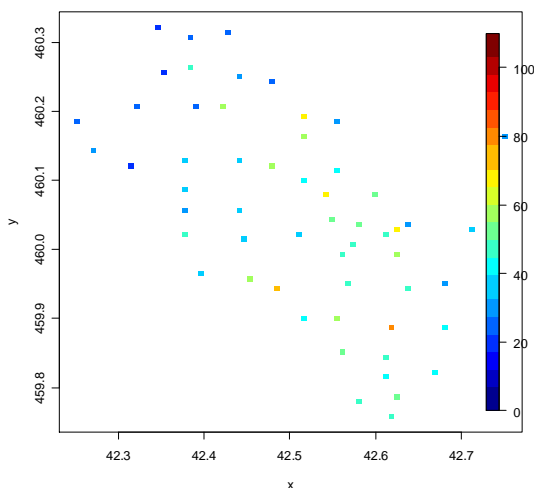


Figura 8. Valors del NO_2 a l'abril en els diferents punts mostrejats en funció de la latitud i la longitud

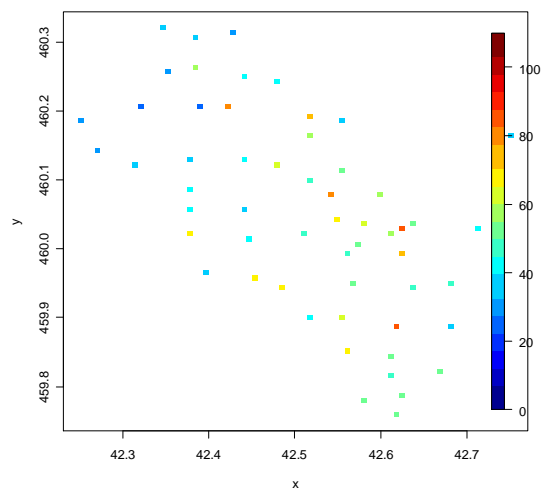


Figura 9. Valors del NO_2 al juny en els diferents punts mostrejats en funció de la latitud i la longitud

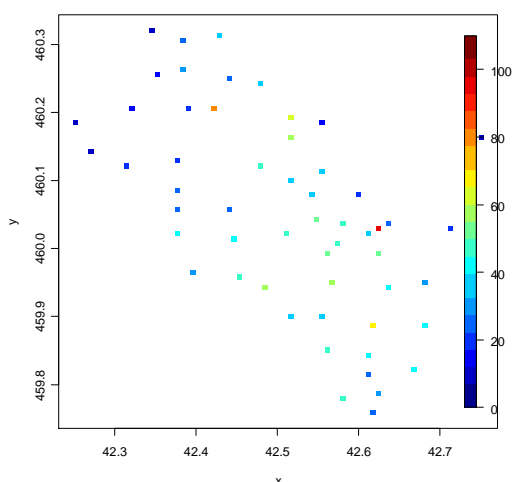


Figura 10. Valors del NO_2 a l'Octubre en els diferents punts mostrejats en funció de la latitud i la longitud

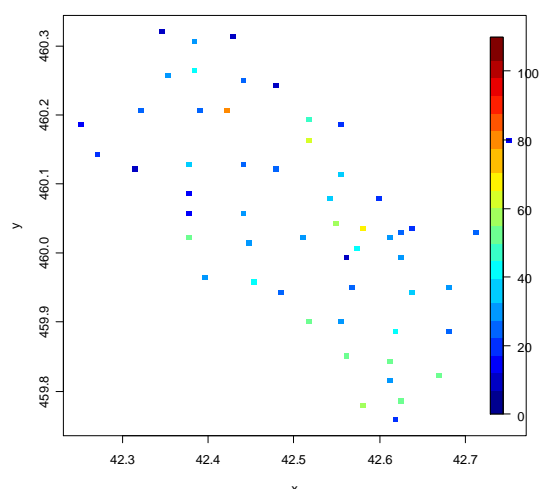


Figura 11. Valors del NO_2 al Març en els diferents punts mostrejats en funció de la latitud i la longitud

La mitjana més alta registrada a les diferents campanyes va ser de $50.1 \mu\text{g}/\text{m}^3$ durant el mes de juny, i la més baixa es va registrar durant el mostreig realitzat al mes de Març del 2006 ($32 \mu\text{g}/\text{m}^3$). La diferencia més gran entre els diferents períodes va ser de $18 \mu\text{g}/\text{m}^3$. Les desviacions dels diferents períodes varen ser molt semblants, al voltant de 15. Amb la idea de tenir una mitjana diària anual, es va fer la mitjana dels quatre períodes. La mitjana de totes les mesures conjuntament va ser de $40.3 \mu\text{g}/\text{m}^3$ i una desviació de 13.8. Únicament la mitjana del més de juny queda per sobre d'aquest valor. Fixant-nos a la mediana ($38.4 \mu\text{g}/\text{m}^3$) observem com són la mediana de juny i la d'abril les que queden per sobre de la mediana dels 4 períodes.

	N	Mitjana	SD	Mediana	Min	Màx.
19-26 Abril 2005	57	42.9	14.8	42.4	18.4	80.4
31-7 Juny 2005	57	50.1	15.4	46.2	25.3	87.5
18-25 Octubre 2005	56	36	18	34.4	7.4	95.4
14-21 Març 2006	56	32	16	28.4	8.4	80.4
Mitjana global	57	40.3	13.8	38.4	18.7	40.3

Taula 5. Promig diària de les mesures de NO_2 recollides als 57 punts de Sabadell durant cadascuna de les campanyes de mostreig

La figura 12 mostra la distribució de les mesures preses per període. S'observa certa simetria en la distància interquartílica en les diferents mesures, així com un comportament asimètric, amb cues més llargues per a valors alts que per a valors baixos. A més, sembla que existeix un patró general definit per un increment de la campanya del mes d'abril a la del mes de juny, un descens del mes de juny al mes d'octubre, i de nou, un increment del mes d'octubre al mes de març.

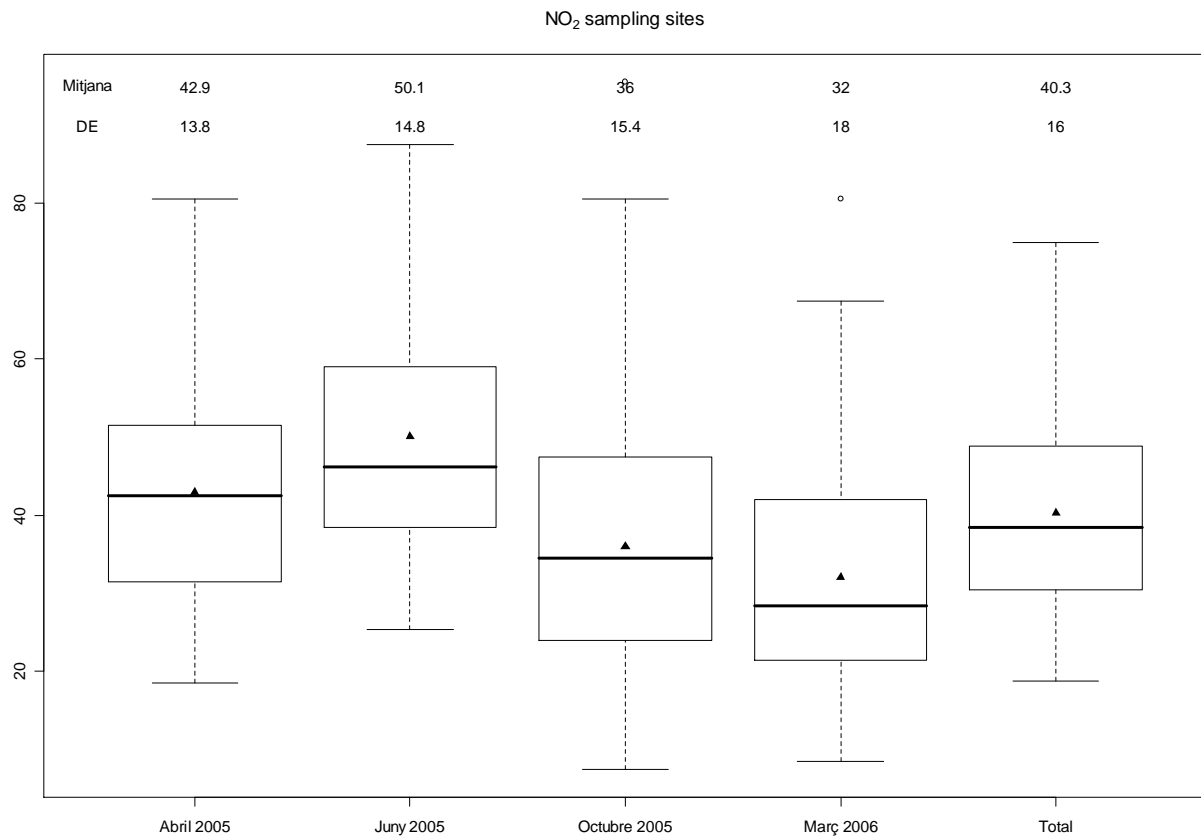


Figura 12. Valors mitjans del NO_2 a les diferents campanyes mostrejades i mitjana global

Al realitzar l'anàlisi de mesures repetides de les dades per campanya de mostreig obtenim un p valor de l'efecte període $<0,001$, pel que rebutgem la hipòtesis nul·la on les mesures són iguals als diferents períodes. Les diferències observades poden ser degudes a les variacions estacionals de l' NO_2 .

Observant els gràfics de dispersió on es creuen les concentracions de NO_2 en les diferents campanyes (figura 13) s'observa com les mesures preses a l'abril i al juny del 2005 són les que mostren una relació lineal més gran, confirmat per la correlació entre les 2 mesures (0.90). Igualment, les mesures de Abril, Octubre, i Juny presenten una relació força lineals, tot i que en menor grau que el primer plot. Les correlacions eren de 0.76 i 0.80 respectivament. D'altra banda, es va detectar que els valors de Març del 2006 són els menys correlacionats amb la resta de mesures. El fet que totes les correlacions fossin altes i positives ens indica que els punts de mostreig de Sabadell queden ordenats per valors de NO_2 de forma similar en les quatre campanyes de mostreig. És per això que la mitjana de les quatre campanyes pot ser una bona mesura dels nivells de NO_2 per cada punt.

	NO ₂ Abril	NO ₂ Juny	NO ₂ Octubre	NO ₂ Març
NO ₂ Abril	1.0			
NO ₂ Juny	0.9	1.0		
NO ₂ Octubre	0.77	0.81	1.0	
NO ₂ Març	0.49	0.53	0.53	1.0

Taula 6. Correlació de les mesures de NO₂ recollides als 57 punts de Sabadell a cada campanya

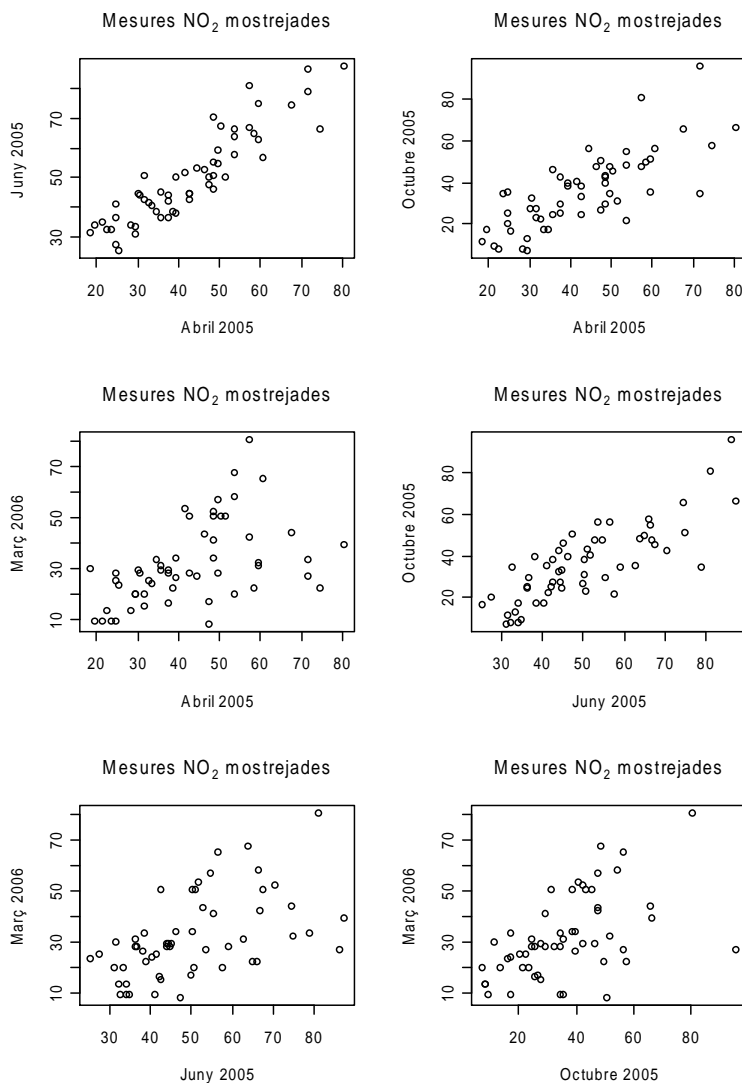


Figura 13. Gràfics de dispersió de les mesures mostrejades de NO₂ en els 57 punts de la ciutat durant les diferents campanyes

7.1.2. Avaluació de les variacions temporals de l'NO₂ recollit a l'estació de trànsit

Per tal d'avaluar quines són les variacions temporals de l'NO₂, es va analitzar la sèrie temporal de les mesures diàries del monitor de trànsit de la ciutat de Sabadell. Per avaluar variació per mes es calcula la mitjana mensual sense tenir en compte l'any de mesura.

S'observa que els valor màxims i mínims mensuals de NO_2 van ser 53.18 $\mu\text{g}/\text{m}^3$ (Juny) i 28.8 $\mu\text{g}/\text{m}^3$ (Agost), i la desviació estàndard mínima era de 15.5 (Juliol) i un màxim de 24.6 (Juny) (Taula 7).

	N	Mitjana	SD	Mediana	Min.	Màx.
Gener	192.0	42.3	24.6	43.4	5.0	144.0
Febrer	169.0	48.8	21.7	50.2	9.0	134.0
Març	202.0	51.3	19.7	49.0	11.3	119.0
Abril	170.0	50.0	18.5	48.0	9.0	97.0
Maig	179.0	50.4	16.6	50.0	6.4	93.7
Juny	146.0	53.2	18.4	55.5	12.8	99.3
Juliol	159.0	48.5	15.5	48.3	9.0	92.3
Agost	164.0	28.1	16.4	27.3	3.1	84.0
Setembre	171.0	50.6	15.4	50.7	17.1	117.1
Octubre	144.0	46.2	19.9	47.9	1.0	100.8
Novembre	153.0	50.9	19.2	50.7	17.5	101.7
Desembre	179.0	46.4	20.7	47.0	1.0	96.0
Total	2028.0	47.2	20.1	47.4	1.0	144.0

Taula 7. Mitjana mensual recollida per l'estació de trànsit de l' NO_2 durant els anys 2001-2007

Així doncs, els valors més alts observats al juny per les nostres mesures concorden amb el patró estacional observat en la sèrie temporal. A la figura 13 podem observar la sèrie temporal diària de NO_2 , on es veu que no hi ha una fluctuació estable. També es mostra la mitjana per mes, per dia de la setmana i per any. S'observa com del gener al juny sembla que la mitjana tingui una mitjana creixent que cau als mesos de juliol, i especialment a l'agost. La resta de mesos tenen un comportament semblant als mesos de gener a juny. Al realitzar l'ajust del model lineal del valor de l' NO_2 respecte al mes s'obté un p valor global <0.0001 .

Pel que fa a la mitjana per dia de la setmana, sembla que hi ha una tendència creixent de dilluns a divendres, i que durant el cap de setmana la mitjana disminueix. Si observem el model ajustat de l' NO_2 respecte el dia de la setmana (taula 8) observem com es confirma el que s'intuïa gràficament. Els coeficients per dissabte i diumenge tenen un valor negatiu, és a dir, que durant aquells dies de la setmana la concentració de NO_2 era més baixa que el dilluns. Per contra, per la resta de dies aquest nivell era més elevat. Per tal d'observar si existia una tendència creixent durant la setmana es va ajustar un model lineal amb el dia de la setmana com a variable continua, començant el dilluns. Es va constatar que el coeficient era positiu i el p valor d'aquest menor a 0.05, pel que podíem afirmar que existia una tendència creixent durant la setmana. En aquesta mateixa taula, es mostra l'ajust de la variable que identifica dia

laborable respecte cap de setmana. Igual que en el model anterior es constata el decrement de la concentració de NO_2 durant el cap de setmana, respecte els dies feiners de la setmana.

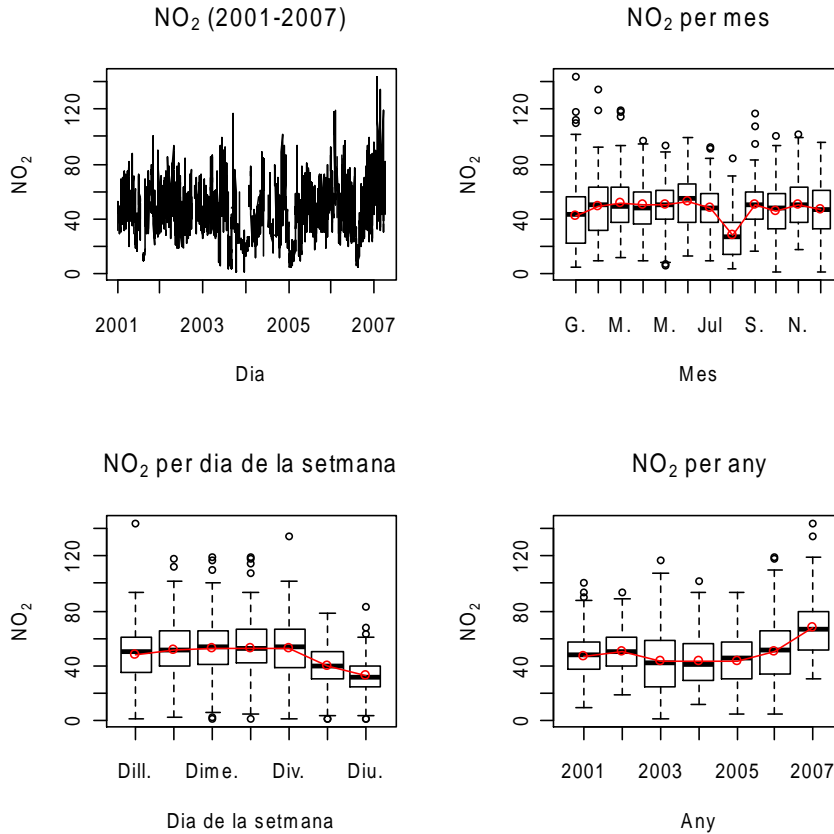


Figura 14. Gràfics temporals de la mesura de NO_2 del captador de trànsit diàriament, per dia de la setmana, mes i any

	Coefficient (SD)	P value	R² ajustada
NO_2 vs. dia de la setmana			
Constant	48.03 (1.11)	< 2e-16	
Dimarts	3.45 (1.57)	0.029	
Dimecres	4.84 (1.55)	0.002	
Dijous	4.88 (1.55)	0.002	
Divendres	5.24 (1.56)	0.001	
Dissabte	-8.32 (1.55)	<0.001	
Diumenge	-15.64 (1.56)	<0.001	13.79%
NO_2 vs. Tipus de dia de la setmana			
Constant	517.29 (0.50)	<0.001	
Cap de setmana	-156.57 (0.92)	<0.001	12.39%

Taula 8. Models del NO_2 dels anys 2001-2007 del captador de trànsit en funció del dia de la setmana

A l'observar les mitjanes en funció de l'any (a la figura 14), s'observa l'existència d'una tendència creixent de l'any 2001 al 2002, del 2002 al 2003 hi ha una disminució, i a partir

d'aquí, s'intueix una tendència creixent fins al 2007 (s'ha de tenir en compte però, que del 2007 només es tenen dades de 3 mesos i per tant la comparació no es vàlida).

Si ens fixem en el model lineal ajustat de l'NO₂ respecte l'any (taula 9) veiem com obtenim la R² ajustada més baixa dels ajustats fins ara amb un 6.66%, és a dir, que l'any és el factor temporal que menys sembla explicar la variabilitat de l'NO₂ dels ajustats fins ara. Respecte als coeficients obtinguts, la diferència màxima entre l'any de referència (2001) i la resta va ser de 20.6, al 2007, i la resta de valors van fluctuar entre -4 i +3, però no sembla que existís una tendència clara en els diferents anys. Al 2002 i al 2006 els coeficients varen ser positius, mentre que els coeficients dels anys 2003 al 2005 varen ser negatius.

	Coeficient (SD)	P value	R² ajustada
Constant	47.22 (1.14)	<0.001	
2002	3.09 (1.58)	0.051	
2003	-4.291 (1.55)	0.006	
2004	-3.60 (1.60)	0.025	6.66%
2005	-3.57 (1.53)	0.020	
2006	3.54 (1.56)	0.024	
2007	20.61 (2.43)	<0.001	

Taula 9. Models del NO₂ dels anys 2001-2007 del captador de trànsit en funció de l'any de mesura

Per últim, i per tal de detectar quin dels factors temporals tenien una influència més gran en la variabilitat de l'NO₂ es va ajustar un model lineal incloent tots tres efectes. A la taula 10 podem observar com l'R² obtinguda amb aquest model s'elevava fins al 30.5%, i com els efectes descrits per separat en els models ajustats anteriorment es mantenen al introduir-los conjuntament al model.

	Coeficient (SD)	P valor	R² ajustada
Constant	40.65 (1.81)	<0.001	30.46%
Febrer	6.43 (1.77)	<0.001	
Març	9.22 (1.69)		
Abril	11.16 (1.79)		
Maig	11.51 (1.77)		
Juny	13.86 (1.87)		
Juliol	9.62 (1.82)		
Agost	-11.08 (1.81)		
Setembre	11.47 (1.79)		
Octubre	8 (1.88)		

Novembre	11.66 (1.84)	
Desembre	7.37 (1.77)	
2002	2.64 (1.37)	
2003	-4.38 (1.34)	
2004	-3.33 (1.39)	
2005	-3.63 (1.32)	
2006	3.45 (1.36)	
2007	22.58 (2.2)	<0.001
Dimarts	3.63 (1.41)	
Dimecres	5.01 (1.4)	
Dijous	4.91 (1.4)	
Divendres	5.23 (1.4)	
Dissabte	-8.39 (1.4)	
Diumenge	-15.59 (1.4)	<0.001

Taula 10. Models del NO₂ dels anys 2001-2007 del captador de trànsit explicat per les variables dia de la setmana, mes i any

Per tal de tenir en compte aquestes dades, i comparar-les amb les recollides en els 57 punts, es van calcular les mitjanes de la sèrie temporal de trànsit en els quatre mateixos períodes de les mostres en els 57 punts (Taula 11). D'aquesta manera es va veure que a les dades del monitor de trànsit, el període amb un valor de NO₂ més baix va ser al mes d'abril, amb un valor de 34.4 µg/m³ i una desviació de 8.5. Per contra, a la mesura corresponent al mes de Juny es va registrar el valor més elevat 59.4 µg/m³ (15.9). Observem doncs com a partir de les dades de l'estació fixa de trànsit hem observat, igual que amb els mostrejos als diferents punts de Sabadell, que la mesura més alta de NO₂ recollida en els diferents períodes i pels 2 mètodes va ser al mes de Juny. En canvi, la mesura més baixa no coincideix.

	N	Mitjana	SD	Mediana	Min	Màx.
19-26 Abril 2005	8	34.4	8.6	34.0	22.0	47.0
31-7 Juny 2005	8	59.4	15.9	58.5	31.0	81.0
18-25 Octubre 2005	8	45.8	7.9	48.5	33.0	56.0
14-21 Març 2006	8	43.6	14.1	47.5	21.0	58.0
Total	32	45.8	14.7	47.5	21.0	81.0

Taula 11. Descriptiva del promig de la mesura de NO₂ de l'estació de trànsit durant els mateixos períodes de les campanyes de recollida dels 57 punts de Sabadell

Podem consirar els quatre valors de l'estació fixa com a valors d'un punt de mostreig addicional a la ciutat. Així, si en un gràfic comparem el patró temporal dels 57 punts amb el punt addicional, observem com el comportament registrat al monitor de trànsit és similar al obtingut a la resta de punts de mostreig (figura 15).

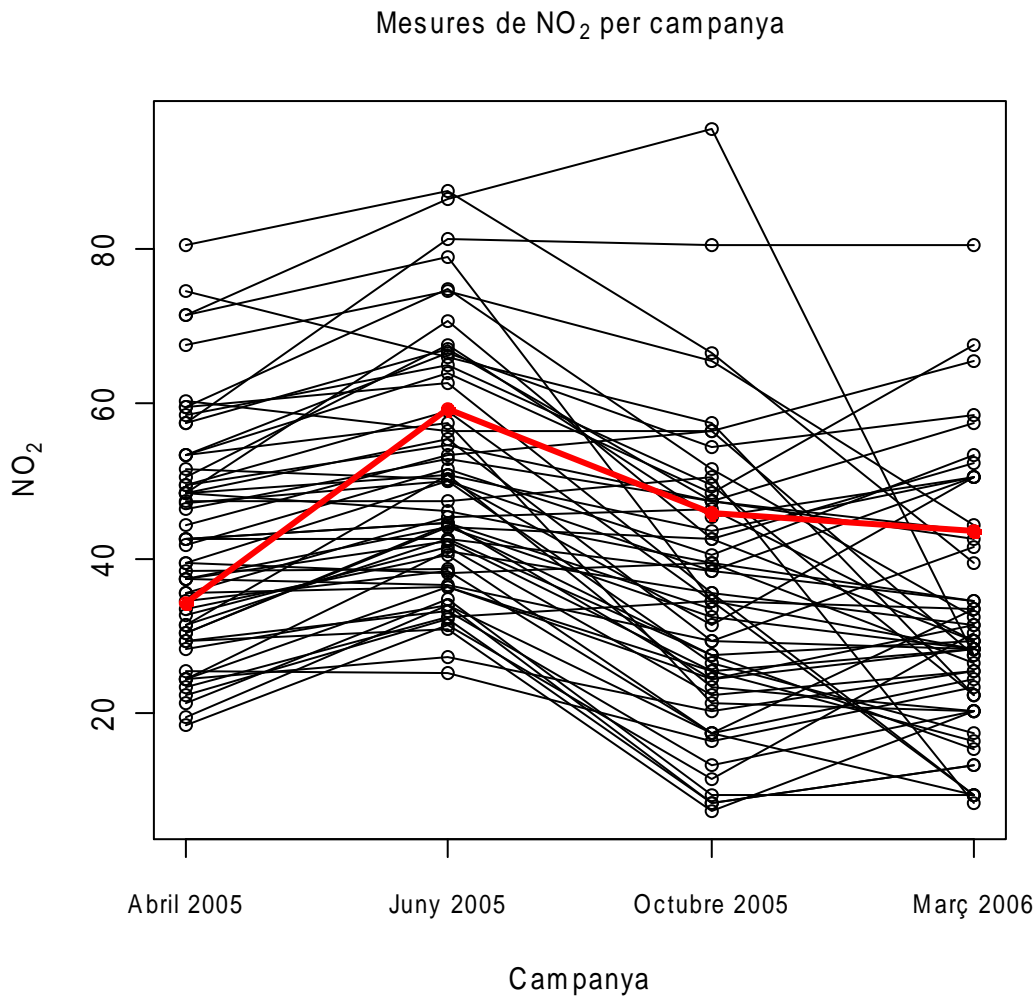


Figura 15. Gràfics de les mesures de NO₂ del captador de trànsit (línia vermella) i de les mesures preses en els 57 punts de la ciutat de Sabadell

Després d'estudiar la relació entre els factors temporals i les dades de NO₂ recollides per l'estació de trànsit podem afirmar l'existència de relació entre la variabilitat d'aquest contaminant en funció del dia de la setmana, el mes i l'any. De la mateixa manera, després de comparar els valors registrats durant les campanyes de mostreig i l'estació de fons sembla com si en general, les mesures preses per l'estació de trànsit fossin més altes que les preses en els punts mostrejats, cosa esperable donat el paper del trànsit en la presència de NO₂. Igualment, observant la figura 15 es detecta un comportament similar pel que fa a la variació de la concentració de l'NO₂ depenent de la campanya. De cara a futurs mostrejos, queda clara la necessitat de continuar realitzant mostrejos a diferents èpoques de l'any, així com la necessitat que aquests incloguin quants més dies de la setmana i mesos millor. D'aquesta manera, el possible efecte temporal, quedaria eliminat al convertir els valors diaris en mitjanes.

En el nostre cas, degut a que la durada entre l'inici i el final de les nostres campanyes de mostreig del diòxid de nitrogen era més gran al període d'embaràs de les dones es va fer servir la informació de l'estació de trànsit per corregir la mesura de NO₂ predita. Per fer-ho es va calcular la mitjana diària durant el període de mostreig i la mitjana, per cada dona, del període d'embaràs. Posteriorment, es va multiplicar la predicció obtinguda pel model lineal per la mitjana diària calculada a l'estació de trànsit en el període d'embaràs de cada dona, i dividir per la mitjana diària del període de mostreig. D'aquesta manera es va corregir la diferència entre el període d'embaràs i la predicció d'exposició durant el període de mostreig.

7.2. Descriptiva univariant de les variables d'interès

7.2.1. Dades de trànsit i geogràfiques per modelar l'NO₂

Les dades que es van recollir referents al trànsit o a la situació geogràfica es van obtenir a partir del Sistema informàtic d'informació ArcGIS o mitjançant el GPS. El primer permet recollir dades referents a:

- a) Topografia (Alçada respecte al nivell del mar)
- b) Densitat de població
- c) Vies de circulació (en funció dels kilòmetres i el volum de trànsit en una distància determinada al punt de mostreig, o de la llar de la dona)
- d) Distància respecte algunes fonts de contaminació (per exemple: pàrkings i benzineres)

En canvi, la localització del punt de mostreig es va determinar via GPS, igual que les adreces de les dones que conformen la nostra cohort.

Les variables que vàrem decidir obtenir a partir del GPS i del programa ArcGis es poden classificar en:

- a) Alçada del punt de mostreig
- b) Distància a possibles fonts de contaminació (benzineres, pàrkings...)
- c) Percentatge d'hectàrees en un cert radi d'un cert tipus de via de trànsit (alta densitat, mitjana i baixa)

Aquestes variables de trànsit, en alguns casos, són molt semblants entre elles. Per exemple, els km de una via amb un alt volum de trànsit en un radi de 50, 75, 100, 150, 250 o 500 metres, o fins i tot, radis entre 50-250. Per tant, el fet que una variable pugui incloure tota

la informació que conté una altra fa que puguin existir correlacions molt altes entre elles. S'ha de tenir en compte que l'elecció dels radis al qual es prenen les mesures venia determinat per estudis previs on s'havia estudiat el mateix.

Per tal de solucionar aquest possible inconvenient a l'hora d'analitzar la relació d'aquestes variables amb l'NO₂ es realitza l'estudi de la relació mitjançant blocs de variables, entenent per bloc de variables a l'agrupació d'aquelles variables que mesuren una mateixa cosa (km de variant la distància únicament (exemple anterior), i també, analitzar tenint en compte les distàncies amb les que s'han mesurat (ex: fer servir totes les variables mesurades dins d'un mateix radi). D'aquesta manera s'intenta evitar la inclusió de col·linealitats dins del model del NO₂, cosa que facilita la interpretació del model.

El procés que es segueix és, primer realitzar l'agrupació de les variables en funció de la seva mesura (km de vies d'alt trànsit, número de pàrkings, número d'edificis entre d'altres). Un cop obtingudes les diferents agrupacions realitzem un anàlisi univariant amb el contaminant i cadascuna de les variables per tal de veure quina d'aquestes presenta una millor R² i per tant, quines expliquen millor el nostre contaminant. Un cop fet això, seleccionem totes aquestes variables i ajustem un model multinomial per poder determinar quines seran les variables que quedaran en el model final. Tot i això, en cas que observem que hi ha molta variació de radis en les mesures, a més de realitzar el procés esmentat, realitzem una nova agrupació de les variables en funció de les distàncies fetes servir per mesurar enlloc del que mesuren les variables, i es repeteix el procés. D'aquesta manera s'intenta tenir en compte tant les col·linealitats per mesura, com per distància.

A continuació mostrem la descriptiva de les diferents variables de trànsit i geogràfiques que es varen recollir pels diferents punts mostrejats, i que posteriorment, van resultar significatives a l'hora de modelar les dades de NO₂ recollides (La resta de variables que es varen fer servir com a possibles explicatives les trobarem a l'annex 3).

Variable	N	Mitjana	SD	Mediana	Mín.	Màx.
Alçada respecte el nivell de l'aigua	57	197.0	30.0	191.6	146.6	260.9
Àrea industrial i Urbana en 500 m (hectàrees)	57	62.4	18.9	67.5	2.8	91.7

Taula 12. Descriptiva del tipus d'àrea en un radi de 500 m al punt de mostreig i alçada respecte el nivell del mar

Variable	N	Freq.
Tipus de carrer	57	

Baix Trànsit	34	59.6%
Mitjà Trànsit	11	19.3%
Alt Trànsit	12	21.1%

Taula 13. Descriptiva de les variables geogràfica tipus de carrer

7.2.2. Covariables de les dones i del nadó

Realitzem una breu descriptiva de les variables sociodemogràfiques que virem considerar d'interès per conèixer la població d'estudi, així com la possibilitat que es convertissin en possibles variables predictores del pes del nadó.

La variable classe social es va crear a partir del reportat al qüestionari general en el camp de feina, a partir de la ocupació de més llarga durada en els últims 10 anys. En cas de no haver treballat, es va fer servir la feina de la parella. La codificació utilitzada és va basar en la International Standard Classification of Occupation, i es va dividir en 6 grups (tot i que degut a la vinculació de les dues primeres categories es va convertir en 5):

- Classe I: Professionals (ex: metge, advocat, comptable).
- Classe II: Ocupacions intermèdies (ex: mestre, infermera).
- Classe III No-manual: treballadors qualificats que treballen en una ocupació no manual (ex: oficinista, secretària).
- Classe III Manual: treballadors qualificats que treballen en una ocupació manual (ex: carnisser, fuster)
- Classe IV: Treballadors parcialment qualificats en una ocupació manual (agricultor, carter, conductor d'autobús)
- Classe V: treballadors no qualificats (ex: netejadors, empleats de fàbrica)
- No classificables: mestresses de casa, estudiants i militars.

Respecte a les característiques socials de les dones de la població de l'estudi observem com aquesta estava formada per dones amb parella estable (97.6%), d'ètnia blanca (96.6%), classe social mitjana-alta (69.8%) i amb un nivell d'estudis mitjà-alt (un 73.8% tenien estudis secundaris o universitaris). Només un 41.2% del total no havien fumats mai, i un 31.8% van fumar durant alguna part o tot l'embaràs. A més, un 62 % feien servir cuina de gas (una de les principals fonts de NO₂ a l'interior de les llars) (Taula 14).

Variable	N	Freq	Freq. Acum
Estat civil			

Variable	N	Freq	Freq. Acum
Casada/parella estable	639	97.56 %	97.56 %
Soltera	10	1.53 %	99.09 %
Separada/Divorciada	6	0.92 %	100.01 %
Total	655	100%	100%
Ètnia			
Blanc	633	96.64 %	96.64 %
Asiàtic	1	0.15 %	96.79 %
Negre	5	0.76 %	97.55 %
Natiu Americà	16	2.44 %	99.99 %
Total	655	100%	
Classe social			
I&II Professionals i Ocupacions intermitges	136	22.0 %	22.0 %
IIINM No Manual	273	44.1 %	66.1 %
IIIM Manual	23	3.7 %	69.8 %
IV/V Treballadors parcialment qualificats I no qualificats	111	17.9 %	87.7 %
No classificables	76	12.3 %	100%
Total	557	100%	100%
Nivell d'estudis			
Primaris	147	26.2 %	26.2 %
Secundaris	239	42.6 %	68.8 %
Universitaris	175	31.19 %	99.99 %
Total	561	100%	100%
Fumadora			
No fuma	251	41.22%	41.22 %
Deixa de fumar abans de l'embaràs	164	26.93 %	68.15 %
Deixa de fumar durant l'embaràs	93	15.27 %	83.42 %
Fuma durant tot l'embaràs	101	16.58 %	100 %
Total	609	100 %	
Cuina de gas			
Si	381	62.05 %	62.05 %
No	233	37.95 %	100%
Total	614	100%	100%

Taula 14. Descriptiva de les característiques demogràfiques i d'hàbits de les dones de l'estudi 1

Respecte a les característiques demogràfiques, destacar que la mitjana d'edat en el moment de l'embaràs era de 31.5 anys, el pes previ a aquest era de 62.6 kg, i l'alçada mitjana de les dones era de 162 cm. L'índex de massa corporal era de 23.7, dins del grup de normopeso.

Variable	N	Mitjana	SD	Mediana	Min	Màx.
Edat de la mare	574	31.5	4.3	31.6	18.2	43.0
Pes de la mare previ embaràs (kg)	611	62.6	12.7	60.0	39.0	143.0
Alçada de la mare prèvia a l'embaràs (cm)	611	162.3	6.1	162.0	146.0	180.0
BMI de la mare previ a l'embaràs (kg/m ²)	611	23.7	4.5	22.7	14.9	53.8

Taula 15. Descriptiva de les característiques demogràfiques de les dones de l'estudi 2

De les dades de l'embaràs i del nadó al néixer cal destacar que en un 56.9% dels casos aquest era el primer embaràs, i l'època de l'any de concepció del nadó va ser molt repartida, cap època destaca respecte les altres. En un 48.7% dels casos el nadó va ser una nena. La mitjana de dies d'embaràs va ser de 278 dies (9.11 mesos) i el pes mitjà del nadó al néixer va ser de 3241 grams. Si ens fixem a la figura 16, on hi ha el boxplot del pes del nadó al néixer, s'hi observa una distribució lleugerament desplaçada a l'esquerra, és a dir, que la cua és més llarga amb valors inferiors a la mediana.

Variable	N	Freq	Freq. Acum
N. d'embarassos previs			
0	320	56.94	56.94 %
1	207	36.83	93.77 %
2	33	5.87	99.64 %
3	1	0.18	99.82 %
5	1	0.18	100%
Total	652	100%	
Època de l'embaràs			
Estiu	174	28.1%	29.2%
Hivern	130	21.0%	50.5%
Primavera	167	27.0%	76.5%
Tardor	148	23.9%	100.0%
Total	619	100%	
Sexe nadó			
Femení	274	48.67 %	48.67 %
Masculí	289	51.33 %	100%
Total	563	100%	100%

Taula 16. Descriptiva de les característiques de l'embaràs i del nadó de les dones de l'estudi 2

* La variable edat gestacional es calcula com la diferència entre la data de l'última regla i la data de naixement.

Variable	N	Mitjana	SD	Mediana	Min	Màx.
Edat gestacional (setmanes)	619	39.7	1.5	39.9	28.4	42.3
Pes al néixer (grams)	616	3241.5	434.4	3282.5	770.0	4480.0

Taula 17. Descriptiva de les característiques de l'embaràs i del nadó de les dones de l'estudi 2

* La variable edat gestacional es calcula com la diferència entre la data de l'última regla i la data de naixement.

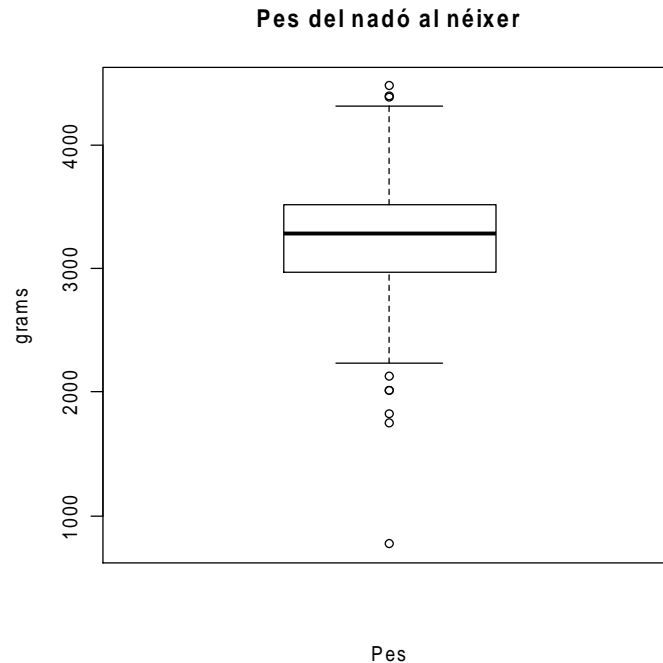


Figura 16. Distribució del pes al néixer dels nadons

7.3. Descriptiva bivariant de les variables explicatives i el pes del nadó al néixer

A continuació varem estudiar com es comportava la variable del pes al néixer en funció dels nivells de cadascuna de les variables candidates a entrar en el model multivariant. Les variables que fem servir en aquests anàlisi són variables que en altres estudis han mostrat un efecte sobre el pes al néixer (per exemple: l'hàbit tabàquic, el sexe del nadó, la classe social), i altres variables recollides a l'estudi (per exemple: tipus de cuina, el gas és una de les principals fonts d'emissió de NO_2 dins de les llars).

A la taula 18 es mostren els valors mitjos del pes al néixer del nadó en funció de les categories de cadascuna de les covariables que es van recollir. La variable ètnia va resultar tenir valors significativament diferents entre els seus nivells tot i que s'ha de tenir en compte el petit nombre d'individus a les categories no blancs. El natiu americà va registrar un valor de 3594.6 grams de mitjana pels 3136.2 dels nadons de ètnia negra, i 3232.3 dels blancs. Pel que fa a la mitjana del pes del nadó en funció del nivell d'estudis de la mare, es va observar que una certa tendència positiva, a més nivell d'estudis, més elevat el pes al néixer, tot i no registrar un p valor <0.05 (límit de significació). Pels nivells de la variable classe social no es van observar diferències significatives, ni l'existència de tendència dins d'aquests nivells. S'ha

de tenir en compte que les variables de classe social i nivell d'estudis estan fortament relacionades entre elles. Una altre variable que en estudis previs han mostrat relació amb el pes del nadó al néixer és l'hàbit tabàquic. En dones fumadores el pes del nadó és en mitjana menor que en dones no fumadores. De la mateixa manera se sap que les nenes pesen menys en el moment del naixement que els nens. En el nostre estudi hem observat que l'hàbit tabàquic mostrava diferències significatives entre els seus nivells. Les dones fumadores varen tenir nadons amb un pes inferior en mitjana a les mares no fumadores. Per últim, l'època de l'embaràs esta prop de la significació estadística. Els nadons concebuts al hivern mostren un pes al néixer més baix que els concebuts en altres èpoques de l'any.

Variable	N	Mitjana (SD)	Mediana	(Mín., Màx.)	p valor
Estat civil					
Casada/parella estable	562	32432.2 (437.7)	3287.5	(770. 4480)	0.50
Soltera	7	3020.7 (455.9)	3250	(2400. 3510)	
Separada/Divorciada	2	3230 (183.9)	3230	(3100. 3360)	
Total	571	3240.4 (437.5)	3280	(770. 4480)	
Ètnia					
Blanc	553	3232.2 (436.5)	3280	(770. 4480)	0.013
Negre	4	3136.2 (72.3)	3122.5	(3070. 3230)	
Natiu americà	14	3594.6 (402.5)	3500	(3000. 4320)	
Total	571	3240.4 (437.5)	3280	(770. 4480)	
Classe social					
I&II Professionals i					
Ocupacions					
intermitges	125	3240.1 (373.9)	3300	(2300. 4085)	0.8
IIINM No Manual	245	3233.7 (427.1)	3275	(1750. 4390)	
IIIM Manual	20	3175.3 (409.1)	3232.5	(2380. 3770)	
IV/V Treballadors parcialment					
qualificats I no qualificats	105	3309.3 (465.2)	3310	(2010. 4480)	Trend: 0.59
No classificables	70	3220.9 (516.2)	3280	(770. 4280)	
Total	565	3245.5 (434.6)	3290	(770. 4480)	
Estudis de la mare					
Primaris	167	3182.8 (472.8)	3180	(770. 4320)	0.13
Secundaris	237	3258.1 (437)	3300	(1830. 4480)	
Universitaris	165	3268.8 (396.4)	3320	(2300. 4390)	
Total	569	3239.1 (437.5)	3280	(770. 4480)	
Edat de la mare					
=<24.5 anys	30	3172.2 (453.2)	3090	(2380. 4075)	0.61
>24.5 i <=30.5 anys	225	3233.3 (434.8)	3295	(1750. 4390)	

>30.5 i <=35.5 anys	228	3256.1 (424.8)	3300 (2010. 4480)	
>35.5 anys	88	3241 (475.2)	3210 (770. 4400)	
Total	571	3240.4 (437.5)	3280 (770. 4480)	Trend: 0.68
Hàbit tabàquic				
No ha fumat mai	235	3257 (432.7)	3300 (1830. 4480)	0.016
Va deixar de fumar abans de l'embaràs	145	3296 (416.3)	3310 (2010. 4280)	
Va deixar de fumar durant l'embaràs	88	3286 (415.1)	3300 (2380. 4400)	
Fumadora	91	3124 (397.5)	3150 (1750. 4075)	
Total	559	3250 (423)	3290 (1750. 4480)	
Cuina de gas				
Si	346	3253 (432.8)	3288 (1830. 4480)	0.93
No	215	3244.2 (406.1)	3295 (1750. 4390)	
Total	561	3249.6 (422.4)	3290 (1750. 4480)	
N. de parts previs				
Cap	323	3220.8 (423.1)	3260 (1750. 4400)	0.18
1 o més	247	3268.7 (456.5)	3300 (770. 4480)	
Total	570	3241.6 (438.1)	3288 (770. 4480)	
Època de l'embaràs				
Hivern	113	3188 (428)	3190 (2010. 4390)	0.06
Primavera	158	3217 (455)	3260 (1750. 4320)	
Estiu	168	3243 (388)	3270 (2015. 4120)	
Tardor	133	3313 (477)	3340 (770. 4480)	
Total	572	3241 (437)	3282 (770. 4480)	
Sexe del nadó				
Masculí	288	3315.8 (401.8)	3340 (1830. 4400)	<0.001
Femení	284	3165.4 (459.3)	3182 (770. 4480)	
Total	572	3241.1 (437.5)	3282 (770. 4480)	

Taula 18. Anàlisi bivariant entre el pes al néixer i la resta de covariables de l'estudi

Capítol 8

Resultats dels models

8.1. Model lineal

8.1.1. Modelatge i predicció del NO₂

El primer pas per realitzar el nostre anàlisi és el de modelar l'NO₂ mesurat en els 57 llocs de Sabadell a partir de les variables descrites anteriorment tant de trànsit com de característiques geogràfiques.

S'ha de tenir en compte que el model del NO₂ és un model de predicció, i per tant el interès es centra únicament en poder obtenir una bona predicció de l'exposició de les dones de l'estudi durant el període d'embaràs. És per això que no interessa tant ni les variables que es fan servir per ajustar al model ni la possible col·linealitat que existeixi entre aquestes. En el model del NO₂ ens fixarem en aquelles variables que resultin ser estadísticament significatives a l'introduir-les al model i que ens incrementin el valor de la R^2 ajustada. D'aquesta manera aconseguirem tenir una bona predicció de l'NO₂. En canvi, en el model del pes es pretén trobar aquells factors que tenen relació amb el pes del nadó al néixer així com utilitzar aquelles covariables que, tot i que no mostrin un efecte directe sobre el pes, pugui presentar un efecte modificador sobre els factors d'interès.

Degut a la no normalitat de les dades de NO₂ mostrejades es farà servir com a variable depenent la transformació logarítmica del NO₂. D'aquesta manera, el model trobat basant-nos en el valor de la R^2 -ajustada (79.3%) i en els millors residus va ser el que incloïa com a variables independents la posició X, la posició Y, la posició X al quadrat (aquest grup de variables ens determina una certa tendència espacial que presentaven les dades), l'alçada del punt de mostreig, el tipus de terreny (terreny urbà i industrial) en un radi de 500 metres, així com el tipus de carrer segons el trànsit (molt trànsit, transit mitjà, o poc trànsit). A la taula 19 es pot observar com a més baixa longitud, més nivell de contaminació (per l'efecte quadràtic de la relació), i a més alta latitud, més baixa. Igualment, a més alçada respecte al nivell del mar, més alta era la concentració de NO₂. I pel que fa al tipus de carrer, com era d'esperar, a més alt volum de trànsit, més concentració de NO₂.

Variable	Coefficient (SD)	IC 95%	P value
Constant	2.43 (0.48)	(1.48, 3.37)	<0.0001
X	1.14 (0.38)	(0.39, 1.88)	0.004
Y	-1.48 (0.34)	(-2.15, -0.81)	<0.0001
X ²	-8.13 (1.6)	(-11.26, -5)	<0.0001
Altitud	0.006 (0.002)	(0.001, 0.011)	0.02
Tipus de carrer			
Mitjà volum de trànsit	0.23 (0.06)	(0.11, 0.35)	0.0004
Alt volum de trànsit	0.48 (0.06)	(0.37, 0.6)	<0.0001

Taula 19. Model de regressió de l'NO₂ als 57 punts mostrejats de Sabadell

Si ens fixem en l'anàlisi dels residus (figura 17) observem com, en el gràfic del residus vs. els ajustats, la distribució d'aquests és aleatòria al voltant del zero, tot i que sembla existir una certa variància no constant, així com l'existència d'alguns valors influents. Si realitzem una breu descriptiva d'aquests possibles valors influents, observem que en un dels quatre casos presenta el valor de NO₂ més alt de tota la població, i un altre cas presenta l'altitud més baixa de la mostra. Si es realitza l'anàlisi sense aquests casos s'observa que el model no varia en excés, per tant, no s'exclouran del model. Al mirar el normal qqplot de la figura 17, s'observa com els residus segueixen la recta de normalitat, tot i la presència, un altre cop d'aquestes 3 dades influents.

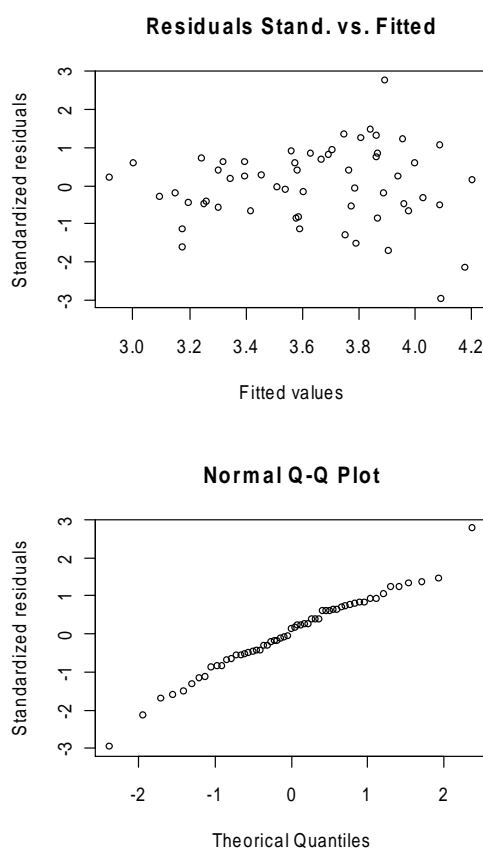


Figura 17. Anàlisi dels residus del model ajustat al promig de les mesures de NO₂ a les 4 campanyes

Donat que el model del NO₂ és un model de predicció, i degut a les limitacions expressades al capítol 2.5, es va calcular l' R² de cross validation per tenir una mesura més objectiva de la capacitat de predicció d'aquest model. També es va estimar l'error de predicció del model. D'aquesta forma vàrem estimar un error de predicció de 66.92, mentre que en el model de predicció sense realitzar bootstrap era de 1.3, i una R² de 64.4%, 15 punts inferior a la estimada mitjançant el model lineal. Així doncs, queda patent que l'R² que s'obté fent servir les mateixes dades per ajustar el model i per avaluar la seva predicció, que és el que habitualment donen molts estudis, està inflat i dóna valors massa elevats.

A partir del model derivat de l'NO₂ (taula 19) es fa la predicció dels nivells de NO₂ als domicilis de les participants.

D'aquesta manera, a la taula 21, observem que la mitjana de la predicció de diòxid de nitrogen obtinguda per les dones era de 35.7 ug/m³ amb una desviació de 9.2. Al boxplot de l'esquerra de la figura 18 es mostra el promig de NO₂ predit a les 611 llars de les dones participants a l'estudi en els diferents períodes de mostreig. S'observa una distribució asimètrica, presenta una cua dreta més llarga que no pas l'esquerra.

Variable	N	Mitjana	SD	Mediana	Min	Màx.
LUR NO ₂	611	35.7	9.2	33.6	20.7	69.3

Taula 20. Predicció de l'exposició de les dones de Sabadell en el període comprès entre les campanyes de mesura

S'ha de tenir en compte però que aquesta predicció es correspon a la mitjana compresa entre els períodes de mostreig. Donat que a nosaltres ens interessa la mitjana d'exposició al NO₂ durant el període d'embaràs cal fer un ajust temporal corregint el fet que la predicció realitzada correspon al promig diari durant la durada de les campanyes. Per fer-ho fem servir les dades de l'estació de trànsit, que al tenir valors diaris, permetrà corregir els nivells. Així, per cada participant, dividim la seva predicció de NO₂ per la mitjana de les dades recollides durant el període de mostreig a l'estació fixa, i multipliquem per la mitjana obtinguda per aquesta mateixa estació durant tot el període d'embaràs. De les 611 dones, únicament 575 varen ser seguides fins a la data de naixement, pel que la correcció del NO₂ la vàrem fer sobre aquestes.

Variable	N	Mitjana	SD	Mediana	Min	Màx.
NO ₂ corregit	575	31.7	9.1	29.6	16.8	69.3

Taula 21. Predicció de l'exposició de les dones de Sabadell durant el període d'embaràs

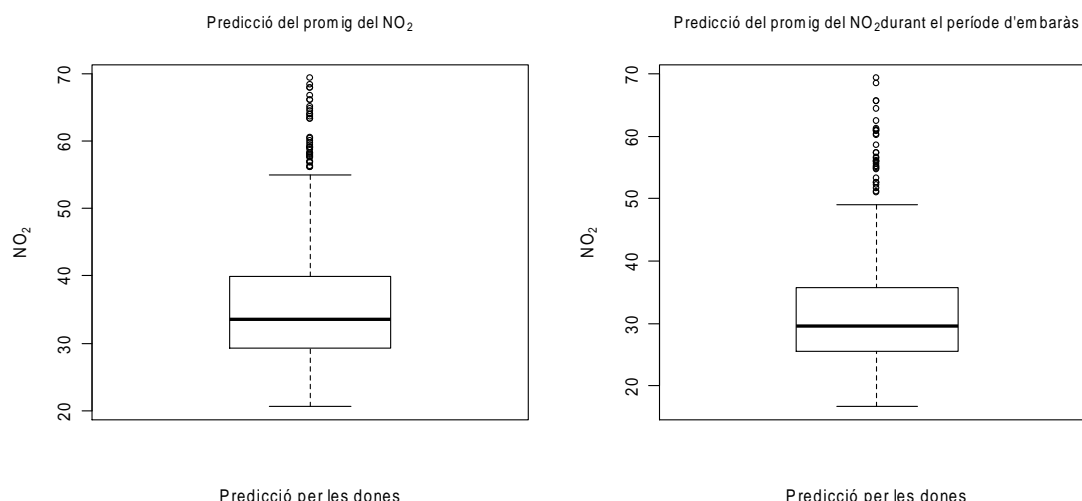


Figura 18. Boxplot de la predicció del NO_2 per les 611 dones residents a la ciutat de Sabadell (esquerra) i Boxplot de la predicció del NO_2 durant el període d'embaràs de les dones residents a la ciutat de Sabadell

El següent pas era el de fer servir la predicció de les dones com una variable explicativa del pes del nadó al néixer. Degut a la presència de missings en el pes al néixer (3 casos) i en algunes de les variables que fèiem servir per a modelar la mostra que fem servir la mostra va quedar en 553 dones.

8.1.2. Descriptiva de les prediccions de l'exposició al NO_2 respecte la resta de covariables

Igual que hem fet a l'apartat 6.3, realitzem una descriptiva bivariant entre l' NO_2 i la resta de variables que podrien intervenir en el model del pes al néixer. D'aquesta manera podrem determinar si poden existir variables confusores, és a dir, variables que estiguin associades tant amb l' NO_2 com amb el pes al néixer i que cal tenir en compte per obtenir estimacions vàlides de l'associació. En aquest cas, aquestes variables, tot i poder no ser significatives en el model del pes al néixer, haurien de ser incloses en el model per tal d'evitar trobar un efecte de la nostra variable d'interès que no es degui a si mateixa sinó a la falta de informació d'altres variables.

A la taula 22 hi ha descrits els valors que pren l' NO_2 en funció dels nivells de cada cofactor. S'observa com l'únic factor que sembla presenta una distribució diferent del NO_2 segons els seus nivells és la variable nivell d'estudis. Per les mares amb un nivell d'estudis primaris o secundaris les mitjanes són de 31.3 i 30.7 $\mu\text{g}/\text{m}^3$ respectivament, mentre que per les mares amb un nivell d'estudis universitaris la mitjana de NO_2 durant el període d'embaràs pren un

valor de $33.7 \mu\text{g}/\text{m}^3$, lleugerament superior als casos anteriors. A diferència de la descriptiva amb el pes al néixer, en aquest cas, la variable estudis de la mare si que va registrar diferències estadísticament significatives. De la mateixa manera, l'estat civil està a prop de la significació estadística, tot i que el fet que les N's dels grups soltera i separada/divorciada siguin menors de 8 fan que el possible efecte estimat es pugui deure a una falta de N en aquests grups. Per la resta de variables no es van observar diferències estadísticament significatives. Tot i això, per les variables que presentaven un gradient a les seves categories, es va realitzar el test de tendència. D'aquesta manera es va observar com la variable d'estudis de la mare com la variable classe social presentaven tendències significatives. Pel que fa al nivell d'estudis de la mare, com ja hem comentat anteriorment, a més nivell d'estudis, més alts són els valors predits del NO_2 durant el període d'embaràs. Pel que fa a la variable classe social, s'observa un gradient positiu, és a dir, a més alta classe social, més alts són els nivells de NO_2 predits durant el període d'embaràs.

Cal destacar que la variable de època de l'embaràs no resulta estadísticament significativa. Això podria contradir el que en el capítol 2.1.2. "Dades de l'estació de trànsit" havíem trobat. S'ha de tenir en compte però, que al haver realitzat un promig dels 9 mesos del període d'embaràs i corregit per l'estació de trànsit, el possible efecte temporal de la mesura ha estat atenuat.

D'altra banda, també es va analitzar la relació del NO_2 amb la resta de variables continues, el pes de la mare abans de l'embaràs, l'alçada, i el IMC, així com l'edat gestacional. Per fer-ho es va constatar la correlació amb les mesures, així com models lineals amb el NO_2 com a variable resposta. En cap cas es va obtenir una correlació significativa ni un p valor significatiu a les regressions. A la figura 19 podem veure com no s'observa una relació clara entre el pes al néixer i l'exposició al diòxid de nitrogen durant el període d'embaràs.

D'aquest manera, les variables que hauríem de considerar d'entrar en el model del pes al néixer, tot i que aquestes no tinguessin un efecte significatiu dins d'aquest, serien les variables de classe social i nivell d'estudis de la mare. Fins i tot podria ser interessant la inclusió de l'època de l'embaràs.

Variable	N	Mitjana (SD)	Mediana (Mín., Màx.)	p valor*
Estat civil				
Casada/parella estable	564	31.6 (9.1)	29.6 (16.8, 69.3)	
Soltera	7	33.6 (6.1)	35.0 (23.1, 40.7)	
Separada/Divorciada	3	43.8 (12.2)	40.4 (33.7, 57.3)	0,06

Variable	N	Mitjana (SD)	Mediana	(Mín., Màx.)	p valor*
Total	574	31.7 (9.1)	29.6	(16.8, 69.3)	
Ètnia					
Blanc	556	31.8 (9.2)	29.6	(16.8, 69.3)	
Negre	4	31.1 (5.0)	31.6	(24.4, 36.7)	
Natiu americà	14	30.2 (6.7)	28.4	(23.3, 43.1)	0,87
Total	574	31.7 (9.1)	29.6	(16.8, 69.3)	
Classe social					
I&II Professionals i Ocupacions					
intermitges	125	33.4 (10.1)	30.0	(17.5, 60.9)	
IIINM No Manual	246	31.7 (9.1)	29.7	(18.6, 65.5)	
IIIM Manual	20	29.9 (6.7)	29.5	(16.8, 44.4)	0,23
IV/V Treballadors parcialment					
qualificats i no qualificats	106	30.6 (8.4)	28.4	(19.0, 69.3)	
No classificables	70	30.9 (8.9)	29.2	(19.0, 68.3)	
Total	567	31.7 (9.2)	29.6	(16.8, 69.3)	Trend: 0.026
Estudis de la mare					
Primaris	168	31.3 (8.8)	28.9	(19.0, 68.3)	
Secundaris	238	30.7 (8.8)	28.7	(16.8, 69.3)	
Universitaris	166	33.7 (9.6)	31.1	(17.5, 64.3)	<0.001
Total	572	31.7 (9.2)	29.6	(16.8, 69.3)	Trend: 0.016
Edat de la mare					
=<24.5 anys	31	29.6 (7.8)	26.1	(20.6, 47.9)	
>24.5 i <=30.5 anys	226	31.2 (9.0)	29.1	(16.8, 69.3)	
>30.5 i <=35.5 anys	228	32.1 (9.2)	29.8	(17.5, 65.6)	
>35.5 anys	89	32.8 (9.7)	30.6	(19.0, 68.3)	0.15
Total	574	31.7 (9.1)	29.6	(16.8, 69.3)	Trend: 0.174
Hàbit tabàquic					
No ha fumat mai	235	31.6 (9.2)	29.8	(18.6, 65.6)	
Va deixar de fumar abans de l'embaràs	146	31.9 (8.6)	29.6	(17.5, 68.3)	
Va deixar de fumar durant l'embaràs	88	31.7 (10.0)	28.9	(18.6, 69.3)	
Fumadora	92	32.1 (9.3)	29.8	(16.8, 61.2)	0,85
Total	561	31.8 (9.2)	29.6	(16.8, 69.3)	
Cuina de gas					
Si	347	31.9 (9.1)	29.7	(16.8, 69.3)	
No	216	31.4 (9.3)	29.2	(17.5, 68.3)	0,41
Total	563	31.7 (9.2)	29.6	(16.8, 69.3)	
N. de parts previs					
Cap	324	32.0 (9.5)	29.8	(17.5, 69.3)	
1 o més	249	31.4 (8.7)	29.3	(16.8, 68.3)	0,72

Variable	N	Mitjana (SD)	Mediana (Mín., Màx.)	p valor*
Total	573	31.7 (9.1)	29.6 (16.8, 69.3)	
Època de l'embaràs				
Hivern	114	31.9 (9.7)	29.0 (20.6, 64.3)	
Primavera	159	32.6 (8.6)	30.3 (17.5, 69.3)	
Estiu	169	31.4 (9.7)	29.1 (18.7, 68.3)	0,12
Tardor	133	30.9 (8.6)	29.7 (16.8, 61.2)	
Total	575	31.7 (9.1)	29.6 (16.8, 69.3)	
Sexe del nadó				
Masculí	290	31.6 (9.0)	29.6 (16.8, 69.3)	
Femení	285	31.8 (9.3)	29.6 (17.5, 68.3)	0,98
Total	575	31.7 (9.1)	29.6 (16.8, 69.3)	

Taula 22. Anàlisi bivariant entre l'NO₂ i la resta de covariables de l'estudi

*Degut a la no normalitat del pes al néixer, es va realitzar el test de kruskal-wallis.

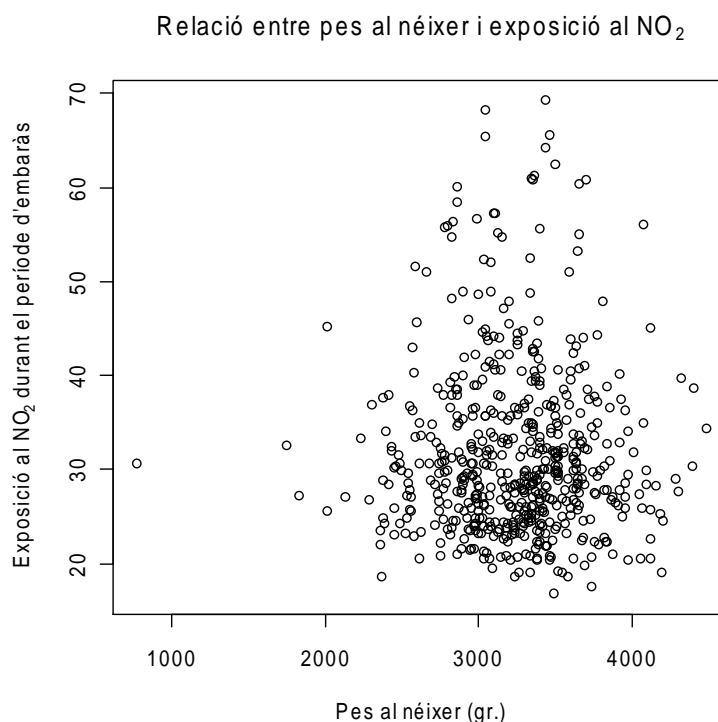


Figura 19. Plot del pes al néixer del nadó i els nivells d'exposició a l'NO₂ de la mare durant el període d'embaràs.

8.1.3. Model final pel pes al néixer

El model final obtingut contenia, a part dels nivells predits de NO₂, les variables sexe, època de l'embaràs, fumadora, nivell d'estudis de la mare, ètnia, número de parts, setmanes de gestació i IMC de la mare abans de l'embaràs. La R² ajustada obtinguda va ser de 22.1%.

A la taula 23 observem els coeficients estimats, el interval de confiança per aquests i el p valor. El fet que el nadó fos una nena, i que la mare fumés durant el període d'embaràs tenia un efecte negatiu en el pes al néixer del nadó (reducció de 130 grams en nenes respecte nens, i de 108.7 grams en dones fumadores durant tot el període d'embaràs). La resta de variables introduïdes al model presentaven un efecte positiu en el pes al néixer, inclosa l'exposició al NO₂. Tot i això, aquesta última variable no va resultar ser significativa (p valor=0.91). Les variables que van registrar coeficients positius més grans varen ser el fet de quedar-se embarassada durant la tardor, ser d'ètnia diferent a la blanca, i el número de setmanes de gestació. En el capítol 8 discutirem els principals resultats obtinguts.

Variable	Coeficient (SD)	IC 95%	P valor
Constant	-1701.34 (480.46)	(-2643.05, -759.64)	<0.0001
Sexe			
Femení	-129.99 (31.85)	(-192.41, -67.56)	<0.0001
Època de l'embaràs			
Primavera	19.44 (47.05)	(-72.78, 111.66)	0.680
Estiu	45.28 (46.11)	(-45.11, 135.66)	0.327
Tardor	141.98 (48.69)	(46.551, 237.418)	0.004
Fumadora			
Deixa de fumar abans de l'embaràs	36.34 (40.38)	(-42.81, 115.49)	0.369
Deixa de fumar durant l'embaràs	45.32 (47.36)	(-47.51, 138.15)	0.339
Fuma durant tot l'embaràs	-108.77 (48.21)	(-203.25, -14.28)	0.024
Nivell d'estudis de la mare			
Secundaris	78.09 (39.23)	(1.19, 154.98)	0.047
Universitaris	83.66 (44.3)	(-3.16, 170.488)	0.059
Ètnia			
Altre a la blanca	233.23 (91.33)	(54.23, 412.23)	0.011
Número de parts			
Un o més d'un	96.25 (32.72)	(32.12, 160.38)	0.003
Setmanes de gestació	114.99 (11.86)	(91.75, 138.24)	<0.0001
NO₂ predit durant el període d'embaràs	0.21 (1.75)	(-3.22, 3.64)	0.906
Índex de massa corporal de la mare			
previ a l'embaràs	11.98 (3.52)	(5.08, 18.89)	<0.0001

Taula 23. Model lineal del pes al néixer fent servir la predicció del NO₂ obtinguda a través del LUR model

8.2. Bootstrap

8.2.1. Modelatge i predicció del NO₂

En aquest apartat repetim les anàlisis realitzades a l'apartat anterior, és a dir, fem servir les prediccions de l'exposició a l'NO₂ per les dones obtingudes anteriorment i estudiem la relació amb el pes al néixer. La diferencia es troba a l'hora d'estimar el coeficient i la desviació de la variable NO₂ en el model de regressió respecte al pes al néixer. En aquest cas, ens ajudem del procés de bootstrap descrit a l'apartat 6.2. per tal de incorporar la incertesa de la predicció del NO₂.

A la taula 24 observem els valors obtinguts pel model lineal ajustat al promig de NO₂ obtingut durant les quatre campanyes de mostreig a la ciutat de Sabadell fent servir el bootstrap. Si tenim en compte els valors obtinguts en el model lineal anterior (taula 19), s'observa que les estimacions des coeficients i els seus intervals de confiança en tots dos casos són molt semblants. Això es degut a que en aquest punt inicial de l'anàlisi mitjançant el bootstrap, no s'incorpora la incertesa de la predicció, donat que encara no hi intervé. Això fa que els valors obtinguts siguin pràcticament iguals als valors del model lineal.

Variable	Coeficient (SD)	Biaix	IC 95%
Constant	2,37 (0,11)	0,11	(1,3, 3,2)
X	1,29 (-0,16)	-0,16	(0,7, 2,2)
Y	-1,27 (-0,17)	-0,17	(-1,7, -0,5)
X ²	-8,2 (-0,06)	-0,06	(-10,6, -5,6)
Altitud	0,006 (0)	-0,0004	(0,002, 0,011)
Tipus de carrer			
Mitjà volum de trànsit	0,31 (-0,08)	-0,08	(0,3, 0,5)
Alt volum de trànsit	0,55 (-0,06)	-0,06	(0,5, 0,7)

Taula 24. Model lineal del valor de NO₂ recollit als 57 punts mostrejats de Sabadell mitjançant bootstrap

8.2.2. Model del pes al néixer

A continuació estimem el model del pes al néixer respecte al NO₂ i la resta de variables fetes servir a l'apartat anterior. En aquest punt és on, a través del remostratge realitzat pel bootstrap, incorporem la incertesa de la predicció de l'NO₂ al model del pes al néixer. Si ens fixem a la taula 25 observem com els coeficients estimats i els seus intervals són molt

semblants per a gairebé totes les variables introduïdes en el model, i registren un valor de biaix molt baix. Fins i tot en el cas de l'efecte del NO₂, on esperàriem un augment dels intervals de confiança degut al fet que incorporem la incertesa sobre les prediccions, no hi ha diferències. El fet que les estimacions del efecte i els seus intervals de les variables diferents al NO₂ sigui molt semblant a les estimacions sense el procés de bootstrap es deu a que en aquestes no presentaven incertesa addicional. En canvi, el fet que l'estimació de l'efecte del NO₂ i el seu interval sigui tan semblant al model lineal sense el procés de bootstrap és, inicialment, inesperat. Pel que sembla, el fet que l'NO₂ no mostri cap tipus d'associació amb el pes al néixer fa que no importi la correcció per la incertesa de l'NO₂. És a dir, si podem dir que el pes al néixer i l'NO₂ són independents, i per tant la seva relació és un núvol de punts aleatori, la relació del pes al néixer vs. l'NO₂ més soroll aleatori (l'error de mesura) continuarà sent igual d'aleatòria. Això queda confirmat per la simulació del capítol 6.2, on també s'ha comprovat que si, per contra, la variable hagués tingut un efecte significatiu, s'hagués observat la incertesa de la predicció en un increment de l'amplitud dels seus intervals de confiança.

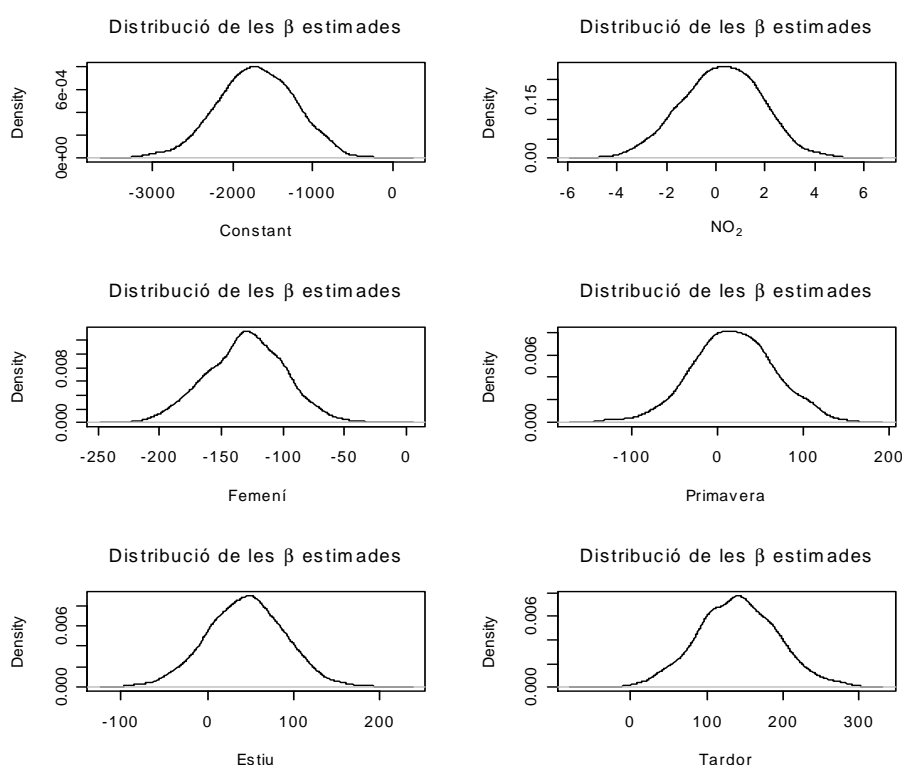
Així doncs, malgrat que en aquest cas la correcció per bootstrap ha resultat ser innecessària, aquest mètode s'aplicarà en el futur en altres estudis de contaminant que tenen un disseny d'estudi similar, que confirmen els resultats asimptòtics sobre la distribució dels coeficients de regressió.

Variable	Coeficient (SD)	Biaix	IC 95%
Constant	-1703.2 (487.3)	-1,9	(-2656.4, -746.1)
Sexe			
Femení	-130 (31.5)	0	(-191.7, -68.2)
Època de l'embaràs			
Primavera	19.4 (48.5)	-0,8	(-74.9, 115.3)
Estiu	45.3 (45.4)	-0,1	(-43.5, 134.4)
Tardor	142.1 (52.1)	-1,7	(41.6, 246)
Fumadora			
Deixa de fumar abans de l'embaràs	36.3 (39)	0,7	(-40.9, 112.1)
Deixa de fumar durant l'embaràs	45.3 (49.6)	-2,1	(-49.8, 144.6)
Fuma durant tot l'embaràs	-108.9 (47.7)	-0,2	(1, -202.2)
Nivell d'estudis de la mare			
Secundaris	78.1 (36.9)	-0,9	(6.6, 151.3)
Universitaris	83.4 (43.3)	-1	(-0.5, 169.3)
Ètnia			

Altres a la blanca	233.3 (99.5)	1,4	(36.8, 427)
Número de parts			
Un o més d'un	96.3 (33.5)	1,2	(29.5, 160.7)
Setmanes de gestació	115 (12)	0,1	(91.4, 138.4)
NO₂ predit durant el període d'embaràs	0.25 (1.6)	0	(-2.9, 3.4)
BMI previ a l'embaràs	12 (3.8)	0	(4.6, 19.4)

Taula 25. Model lineal del pes al néixer incorporant la incertesa de la predicció del NO₂ mitjançant bootstrap

A la figura 20 observem els diferents valors que s'obtenen en cadascun dels models realitzats sobre cada bootstrap per cadascun dels coeficients del model lineal. En gairebé totes elles sembla que s'observa una distribució normal dels valors que prenen.



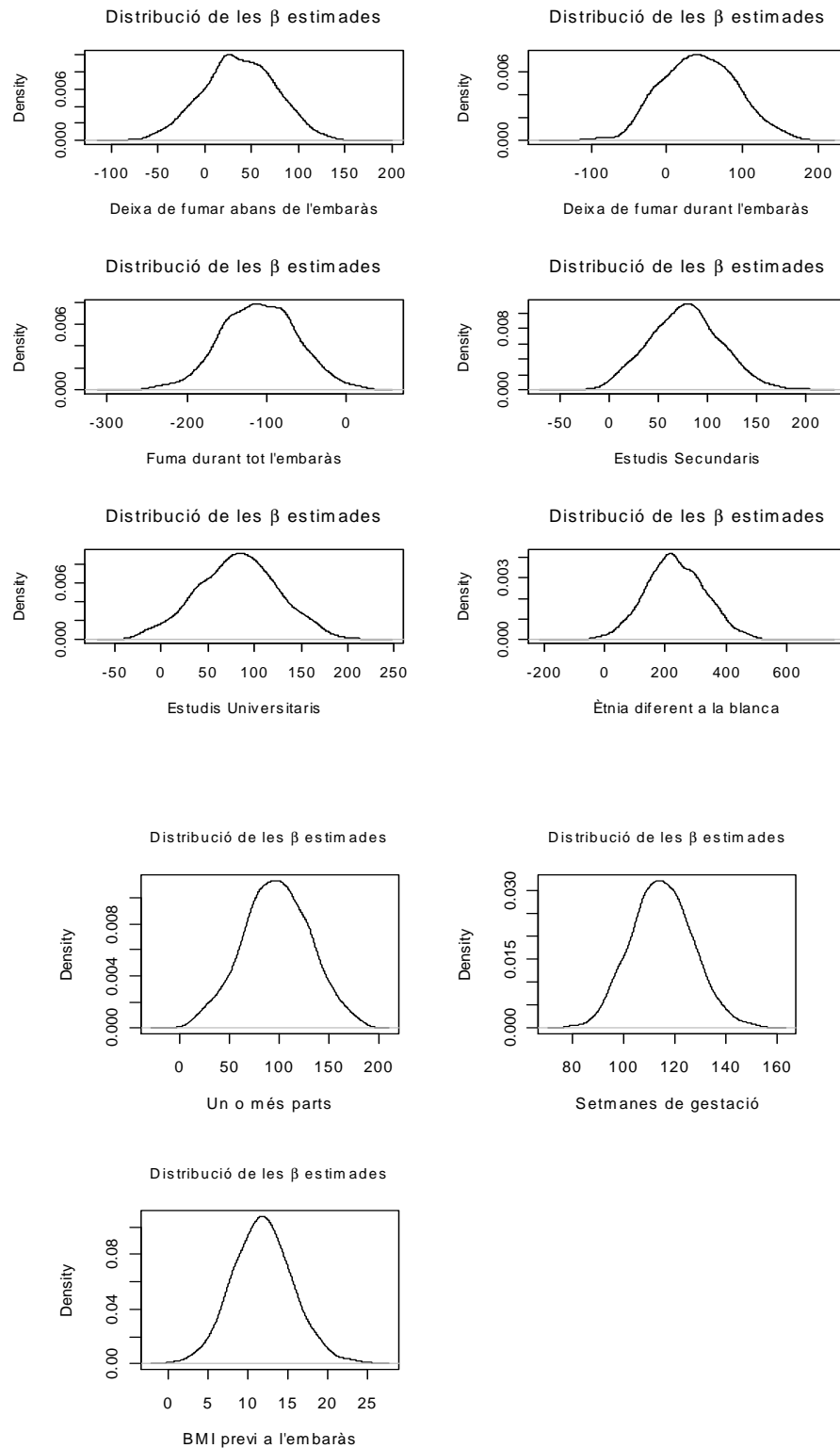


Figura 20. Histograma de l'estimació dels coeficients del model del pes del nadó al néixer

Capítol 9

Conclusions

En aquest projecte hem vist un mètode a partir del qual en epidemiologia s'estima l'exposició al NO₂ de les persones durant un determinat període i el seu efecte en una determinada variable de salut. S'han incorporat millores en l'anàlisi i s'han comparat amb possibles alternatives. Tot i això, no s'han trobat un efecte de l'NO₂ en el pes del nadó, i aquest fet també ha comportat que algunes de les correccions implementades no tinguessin cap efecte. El fet que no s'hagi trobat associació pot ser degut a que l'associació realment no existeixi, o a diverses limitacions de l'estudi. En el següent capítol comentarem algunes d'aquestes limitacions i possibles solucions de cara a futurs estudis.

Fixant-nos en els resultats obtinguts, primer de tot comentem el model d'ajust i de predicció del diòxid de nitrogen. Hem observat com, fent servir el model lineal i amb l'ajut d'un número reduït de variables (4), s'obtenia un coeficient de determinació alt (79.3%), indicant que amb aquestes variables es poden obtenir prediccions més o menys fiables dels valors mitjans de NO₂ en un punt de la ciutat. Tot i això, també hem observat com, en el moment en que consideràvem el model del NO₂ com a model de predicció (com fem servir en aquest projecte) i a través de la metodologia de cross validation, aquest coeficient disminuïa fins al 64.4%, reflectint un valor més objectiu del poder predictiu del model. Aquest fet és un punt que habitualment, en epidemiologia, no es té en compte i es tendeix a oblidar tot assumint que l'R² del model d'ajust serveix per mesurar la capacitat predictiva del model.

Si entrem més al detall de les variables utilitzades per ajustar el model, observem com, a la ciutat de Sabadell existia un patró espacial determinat per la latitud i la longitud, així com l'alçada respecte al nivell del mar. La quarta variable que va resultar significativa en el model va ser el tipus de carrer en el qual estava el punt de mostreig. S'ha de tenir en compte que l'NO₂ pot presentar molta variabilitat dins d'una mateixa ciutat. És per això que les components espacials poden tenir molta importància a l'hora d'explicar el seu comportament. Aquest fet es va observar també a la descriptiva de les dades en el capítol 7. De la mateixa manera, el diòxid de nitrogen està molt vinculat al trànsit, d'aquí la presència de la variable de tipus de carrer en funció del trànsit present. Segurament, el fet que la ciutat de Sabadell sigui una àrea de terreny no molt extensa i amb un tipus de vies de circulació molt similar dins la pròpia ciutat,

fa que altres variables que en altres ciutats podrien tenir importància en l'ajust del model del NO₂, en aquest cas, no fos així.

Arribats a aquest punt, s'ha de recordar que durant el modelatge, predicció i posterior ajust del model del pes l'NO₂ sempre s'ha tractat com una mitjana diària del període d'embaràs, motiu pel qual el factor temporal quedava eliminat. En el següent capítol comentarem els problemes de fer servir la mesura d'aquesta manera.

Posteriorment, un cop realitzat l'ajust del model del NO₂ als punts de mostreig, es va realitzar la predicció per les dones de l'estudi fent servir aquest predicció corregida pel promig del NO₂ durant el període d'embaràs. Aquesta mesura de correcció es va obtenir de l'estació de trànsit. Tot i això, no semblaven existir diferències entre els valors mitjans obtinguts sense realitzar la correcció, i un cop feta la correcció.

Pel que fa al model del pes al néixer es va realitzar l'ajust mitjançant un model lineal multivariant, així com un model lineal multivariant realitzant bootstrap. En tots dos casos les variables utilitzades varen ser les mateixes. Totes elles varen resultar significatives a excepció del NO₂, però la vàrem deixar dins del model perquè era la variable d'interès de l'estudi. En tots dos casos els valors obtinguts, tant de l'efecte estimat com del seu interval de confiança, van ser molt iguals, i en cap cas, si una variable va resultar significativa en el model lineal multivariant va deixar de ser-ho al realitzar el bootstrap.

Pel que fa a la magnitud del efectes estimats, en ambdós casos el fet de ser nena respecte ser nen implicava un pes mitjà inferior en 130 grams. En el mateix sentit, el fet de ser un nadó de mare fumadora implicava un pes mitjà menor en 109 grams respecte una mare que no hagués fumat mai. Pel que fa a les variables que van registrar un efecte positiu sobre el pes del nadó, destacar que el fet d'haver estat concebut a la tardor respecte a haver-ho estat a l'hivern feia que el pes del nadó, en mitjana, s'incrementés en 142 grams. Aquest efecte podria estar amagant un possible efecte de la dieta, de l'activitat física o de la temperatura. Depenent de l'època de l'any de l'embaràs tant la dieta com l'activitat física varia. Tot i això, el fet de no tenir una bona informació referent a l'activitat física durant el període d'embaràs i la complexitat de la informació de la dieta durant aquest període no ens ha permès comprovar aquesta hipòtesi. Una altra variable que va implicar un increment significatiu del pes al néixer va ser el nivell d'estudis. El fet de tenir estudis secundaris respecte a tenir-ne primaris o no tenir-ne feia que el pes del nadó en mitjana fos al voltant de 80 grams superior. Això podria ser degut al fet que una dona amb un nivell d'educació més alt acostuma a tenir una vida més activa, i és més conscient d'una dieta equilibrada. Dues altres variables que estan

relacionades amb un increment del pes en mitjana similar al del nivell d'estudis de la mare eren el fet d'haver tingut parts previs i el número de setmanes de gestació. Les dones que pertanyien al grup amb un o més parts previs es relacionava amb una mitjana de pes superior a les mares sense part previ en 96 grams. Pel que fa a la setmana de gestació, per cada setmana de gestació el increment de pes estimat era de 115 grams. En estudis previs la relació entre les setmanes de gestació i el pes del nadó al néixer era directa, a més setmanes de gestació major pes. De la mateixa manera, per cada unitat d'increment de l'índex de massa corporal de la mare previ a l'embaràs el pes del nadó s'incrementava en 12 grams. Per últim, pertànyer a una ètnia diferent a la blanca tenia un efecte positiu sobre el pes del nadó al néixer, amb un increment mitjà de 233 grams. Tot i això, el fet que el grup de dones amb ètnia diferent a la blanca ós petit fa que tot i que l'efecte estimat és gran, es tinguin certs dubtes sobre la magnitud d'aquest. Així doncs, sembla que de tots els factors introduïts en el model, els únics que van registrar un efecte negatiu respecte el pes al néixer varen ser el sexe femení del nadó, i el fet de tenir una mare fumador. La resta de variables introduïdes al model van registrar un efecte positiu sobre el pes del nadó al néixer.

L'única variable que vàrem deixar al model sense que ós significativa va ser l' NO_2 , la de més interès en el nostre estudi. Això podria ser degut a que, realment no existís cap relació entre el pes al néixer i l'exposició a l' NO_2 durant el període d'embaràs, o bé, a que la mesura de la que disposem, tot i provenir d'un model de predicció amb una R^2 bastant gran (64.4%) no recollia correctament l'exposició durant el període d'embaràs. S'ha de tenir en compte que el que s'ha estimat és el promig de NO_2 a les ubicacions de les residències de la casa, i que per tant, tot i que es pressuposa que gran part del temps la dona estarà prop del punt estimat, hi han moltes altres condicions que no s'han tingut en compte. En el següent apartat discutirem amb més profunditat, i donarem possibles solucions.

El fet que el promig d'exposició durant el període d'embaràs a l' NO_2 no mostrés relació amb el pes al néixer ens va impedir de determinar quin grau d'incertesa presentava l'efecte estimat al fer servir, enlloc de la mesura real, la predicció realitzada a través del LUR model. Per tant, no es va poder determinar el grau d'error que es cometia al obviar el fet que la mesura era predita i no observada, ni la utilitat de fer el bootstrap a més de realitzar el model lineal multivariant.

Capítol 10

Limitacions / Treball futur

La principal limitació amb la que ens hem trobat en aquest estudi, i la que ha promogut el mateix ha estat la dificultat de poder tenir una mesura d'exposició al diòxid de nitrogen per a cada dona de l'estudi. El fet que prendre aquesta mesura sigui excessivament cara impossibilita aquest fet, i per tant, obliga a buscar solucions. En el nostre cas, disposàvem de quatre campanyes de captació del NO₂, de 7 dies de durada cadascuna, i de les dades de 7 anys d'una estació de trànsit d'aquest mateix contaminant. Això permet obtenir una bona estimació de l'exposició exterior al contaminant, tot i que en aquest cas, la manca de variables com direcció del vent dels carrers, o l'existència de corredors de vent ("street canyons", degut a que per la construcció i el trànsit es formen una mena de canons de contaminació") han evitat l'obtenció d'una mesura millor. Tot i això, l'exposició individual és molt més complexa que no pas la mesura exterior ja que té moltes més components i factors que hi influeixen. Mitjançant el bootstrap hem provat de millorar aquesta mesura, al considerar, a diferència de molts estudis, aquesta mesura com a predicció, i no com a valor real. El fet que al relacionar la nostra mesura amb el pes al néixer aquesta hagi sortit no significativa, ens ha impedit de veure la diferència entre fer servir i no fer servir el bootstrap, i la utilitat d'aquesta eina.

Per tal d'evitar aquest problema, i com ja s'ha fet en altres estudis, una opció podria ser, a més de realitzar la predicció del NO₂ durant el període d'embaràs mitjançant el LUR model, el poder disposar de mostres personals de una submostra de la població en estudi. Aquestes mostres podrien ser preses a les mateixes dones, mesures dins i fora de la llar i, en cas que treballin, dins i fora de la feina. Aquestes noves mesures, combinades amb el temps transcorregut a cadascun dels diferents llocs podria permetre'ns realitzar una millor estimació del valor mitjà de l'exposició al NO₂ durant el període d'embaràs. Tot i això, no n'hi hauria prou en recollir aquestes noves dades, sinó que també s'hauria de pensar una manera d'integrar totes aquestes mesures a través d'un o diversos models, i quina seria la millor tècnica estadística per fer-ho. Una possible manera de dur a terme aquest anàlisi seria el de modelar les mesures de dins i fora de la llar, i de dins i fora de la feina, i fer-ne la predicció per totes de dones de l'estudi mitjançant variables recollides per totes les dones i que s'associen amb totes aquestes mesures. Un cop modelat aquestes, fer-ne la predicció, i utilitzar-la per predir la mesura personal de NO₂. D'aquest manera, sembla que una metodologia jeràrquica en el

sentit dels models tindria sentit, i donat que parlem de prediccions, hauríem d'utilitzar alguna eina estadística o metodologia per considerar aquest fet. Una possible solució seria fer servir el bootstrap per tal d'incorporar la incertesa de les prediccions, tal i com hem fet en aquest estudi. Una altre possible solució, que també hem comentat en el projecte, seria el fer servir models bayesians. Fer servir aquesta metodologia permetria modelar simultàniament tots els models implementats, amb la qual cosa, la incertesa que podria haver-hi pel fet de fer servir variables predites enlloc de variables mesurades quedaria incorporada directament en el model. Recordar que aquesta metodologia treballa amb distribucions, i no amb valors puntuals, i d'aquí que incorpori directament la possible incertesa de les prediccions.

Bibliografia

Bates DV. A half century later: recollections of the London fog. *Environ Health Perspect* 2002; 110(12):A735.

Bell ML, Davis DL. Reassessment of the lethal London fog of 1952: novel indicators of acute and chronic consequences of acute exposure to air pollution. *Environ Health Perspect* 2001; 109 Suppl 3:389-394.

Nieuwenhuijsen MJ. Introduction to exposure assessment. In: Nieuwenhuijsen MJ, editor. *Exposure assessment in occupational and environmental epidemiology*. Oxford: Oxford University Press, 2004: 3-19.

de Hartog JJ, Hoek G, Mirme A, Tuch T, Kos GP, Ten Brink HM et al. Relationship between different size classes of particulate matter and meteorology in three European cities. *J Environ Monit* 2005; 7(4):302-310.

Hunter PR, Bickerstaff K, Davies MA. Potential sources of bias in the use of individual's recall of the frequency of exposure to air pollution for use in exposure assessment in epidemiological studies: a cross-sectional survey. *Environ Health* 2004; 3(1):3.

Gustafson P, Le ND, Vallee M. A Bayesian approach to case-control studies with errors in covariables. *Biostatistics*. 2002 Jun;3(2):229-43.

John Molitor, Michael Jerrett, Chih-Chieh Chang, Nuoo-Ting Molitor, Jim Gauderman, Kiros Berhane, Rob McConnell, Fred Lurmann, Jun Wu, Arthur Winer, and Duncan Thomas. Assessing Uncertainty in Spatial Exposure Models for Air Pollution Health Effects Assessment. *Environmental Health Perspectives*, Volume 115; Number 8; August 2007

Jordi Sunyer, Carme Puig, Maties Torrent, Oscar Garcia-Algar, Ignasi Calicó, Laura Muñoz-Ortiz, Maria Barnes, Paul Cullinan on behalf of the AMICS study. Nitrogen dioxide is not associated with respiratory infection during the first year of life. *International Journal of Epidemiology*. 2004. PMID: 15075156

A. Esplugues, R. Fernández-Patier, I. Aguilera, C. Iñiguez, S. García Dos Santos, A. Aguirre Alfaro, M. Lacasaña, M. Estarlich, J. O. Grimalt, M. Fernández, M. Rebagliato, M. Salah, A. Tardón, M. Torrent, M.D. Martínez, N. Ribás-Fitó, J. Sunyer, F. Ballester. Exposición a contaminantes atmosféricos durante el embarazo y desarrollo prenatal y neonatal: protocolo de investigación en el proyecto INMA (Infancia y Medio Ambiente).

Aguilera I, Sunyer J, Fernández-Patier R, Hoek G, Aguirre-Alfaro A, Meliefste K, Bomboi-Mingarro MT, Nieuwenhuijsen MJ, Herce-Garraleta D, Brunekreef B. Estimation of outdoor NOx, NO2 and BTEX exposure in a cohort of pregnant women using Land Use Regression modeling. *Environ Sci Technol* 2008; 42 (3): 815-21.

Brunekreef, B.; Holgate, S. T. Air pollution and health. *Lancet*. 2002, 360, 1233-1242.

Association between GIS-based exposure to urban air pollution during pregnancy and birth weight in the INMA Sabadell Cohort. Aguilera I, Guxens M, Garcia-Esteban R, Corbella T, Nieuwenhuijsen MJ, Foradada CM, Sunyer J. *Environ Health Perspect*. 2009 Aug;117(8):1322-7. Epub 2009 Apr 13.

- Nieuwenhuijsen M, Martinez D, Grellier J, Bennett J, Best N, Iszatt N, Vrijheid M, Toledano M. *Chlorination Disinfection By-Products in Drinking Water and Congenital Anomalies: Review and Meta-Analyses*. Environ Health Perspect 2009; 117(10): 1486-1493.
- Dugas J, Nieuwenhuijsen M, Martinez D, Iszatt N, Nelson P, Elliott P. Use of biocides and insect repellents and risk of hypospadias. Occup Environ Med 2009
- Vrijheid M, Martinez D, Forns J, Guxens M, Julvez J, Ferrer M, Sunyer J. Prenatal Exposure to Cell Phone Use and Neurodevelopment at 14 Months. *Epidemiology 2010: (en Premsa)*.
- Olsen J. Prenatal exposures and long-term health effects. Epidemiol Rev 2000; 22(1):76-81.
- Schwartz J. Air pollution and children's health. Pediatrics 2004; 113(4 Suppl):1037-1043.
- Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A cohort study of traffic-related air pollution impacts on birth outcomes. Environ Health Perspect 2008; 116(5):680-686.
- Slama R, Morgenstern V, Cyrus J, Zutavern A, Herbarth O, Wichmann HE et al. Traffic-related atmospheric pollutants levels during pregnancy and offspring's term birth weight: a study relying on a land-use regression exposure model. Environ Health Perspect 2007; 115(9):1283-1292.
- Perera FP, Rauh V, Tsai WY, Kinney P, Camann D, Barr D et al. Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. Environ Health Perspect 2003; 111(2):201-205.
- Parker JD, Woodruff TJ. Influences of study design and location on the relationship between particulate matter air pollution and birthweight. Paediatr Perinat Epidemiol 2008; 22(3):214-227.
- Gauderman WJ, Avol E, Lurmann F, Kuenzli N, Gilliland F, Peters J et al. Childhood asthma and exposure to traffic and nitrogen dioxide. Epidemiology 2005; 16(6):737-743
- Nyberg F, Gustavsson P, Jarup L, Bellander T, Berglind N, Jakobsson R et al. Urban air pollution and lung cancer in Stockholm. Epidemiology 2000; 11(5):487-495.
- Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 2008; 42(33):7561-7578.
- Lebrete E, Briggs D, Van Reeuwijk H, Fischer P, Smallbone K, Harssema H et al. Small area variations in ambient NO₂ concentrations in four European areas. Atmos Environ 2000; 34:177-185
- Henderson, S. B.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. Environ. Sci. Technol. 2007, 41, 2422-2428.
- Aguilera I, Sunyer J, Fernández-Patier R, Hoek G, Aguirre-Alfaro A, Meliefste K, et al. 2008. Estimation of outdoor NO_x, NO₂, and BTEX exposure in a cohort of pregnant women using land use regression modeling. Environ Sci Technol 42:815-821.
- An Introduction to the Bootstrap, Vol. 57 by Bradley Efron, R.J. Tibshirani (1994)
- Rousson V, Gosoni N F. An R-square coefficient based on final prediction error. ScienceDirect. Statistical Methodology. doi:10.1016/jostatmet.2006.11.04
- "Bayesian Data Analysis," by Gelman, Carlin, Stern, and Rubin (1995, 2004)

Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition
Raymond J. Carroll, David Ruppert, Leonard A. Stefanski and Ciprian Crainiceanu (2006)

Efecto de la exposición a contaminación atmosférica durante el embarazo sobre el crecimiento fetal. Inmaculada Aguilera Jiménez (TESI DOCTORAL UPF)

Annexes

12.1. Annexe 1: Exemple de Qüestionaris

CUESTIONARIO GENERAL: VISITA DE INCLUSIÓN (V₁)

DATOS PERSONALES DE IDENTIFICACIÓN

NOMBRE :
 APELLIDOS:
 FECHA DE NACIMIENTO: DNI:
 TARJETA SANITARIA (SIP):.....
 CENTRO DE ATENCIÓN PRIMARIA al que pertenece:
 NÚMERO DE SEGURIDAD SOCIAL (NSS):
 N° HISTORIA CLINICA DE LA FE:

DIRECCIÓN ACTUAL

Calle N° Piso Puerta
 Municipio CP
 Teléfono casa: Teléfono trabajo:
 Teléfono(s) móvil(es):
 e-mail:
 N° Teléfono de posibles contactos (familiares, amigos)

Parentesco: Teléfono: Móvil:

S. DATOS SOCIODEMOGRÁFICOS

S1. Lugar de nacimiento (municipio)	S2. Provincia/Región	S3. País	S1 <input type="text"/> <input type="text"/> S2 <input type="text"/> <input type="text"/> S3 <input type="text"/> <input type="text"/> S4 <input type="text"/> <input type="text"/>
S4. Grupo étnico: (No realizar esta pregunta, contestar por observación y en caso de duda verificarlo)			
1 Blanco 2 Asiático 3 Negro 4 Árabe (incluidos Norte de África y Oriente Medio) 5 Gitano			

Amerindios
Otros, especificar
9 Ns/Nc

S5. ¿Cuántos años lleva viviendo en su municipio actual? Nº años Nº meses

S5

S6. ¿Desde cuando vive en su casa actual? (mes, año)

S6

S7. Si menos de 10 años, rellenar el cuadro siguiente (*hacia atrás empezando por el último domicilio hasta cubrir un máximo de 10 años*)

Calle, nº	Municipio/ Provincia/País	Desde (mes, año)	Hasta (mes, año)
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

S8. ¿En qué zona del municipio vive usted?

- 1 Casco urbano o barrio periférico
- 2 Urbanización
- Campo

9 Ns/Nc

S9. ¿Cuál es su estado civil?

- 1 Casada o pareja estable
- 2 Soltera
- 3 Separada o divorciada
- 4 Viuda
- 5 Otros, especificar
- 9 Ns/Nc

S8

S9

S10. ¿Cuál de las siguientes formas define mejor la situación en la que vive usted?

- 1 Vive con el padre del bebé que espera
- 2 Vive con otra pareja
- 3 Vive con sus padres
- 4 Vive sola
- 5 Otros, especificar

S10

S11. ¿Qué nivel de estudios ha finalizado? (*no leer las opciones*)

- 1 No sabe leer ni escribir
- 2 Sin estudios o estudios primarios incompletos
- 3 Estudios primarios (EGB, primaria, ESO)
- 4 Estudios secundarios (BUP, COU, FP)
- 5 Estudios universitarios
- 6 Otros (especificar).....

S11

L. HISTORIA LABORAL DE LA MADRE

L1. ¿Cuál es su situación laboral actual?

- Trabajadora
Parada

SI LA RESPUESTA ES 1 ó 4 PASAR A LA PREGUNTA L3

L1

3	Estudiante	
4	Baja laboral Por:	
5	Ama de casa	
6	Otras (especificar):	
L2. (Sólo si no está trabajando actualmente) ¿Ha trabajado alguna vez durante los últimos 10 años?		L2
1	Sí	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> SI LA RESPUESTA ES NO o Ns/Nc PASAR A LA PREGUNTA P1 </div>
2	No	
9	Ns/Nc	
L3. En relación con su trabajo actual y/o trabajos anteriores realizados durante los últimos 10 años, podría indicarnos (empezar por trabajo actual o último y continuar hacia atrás hasta cubrir los 10 años si ha habido otros trabajos previos)		
Periodo de contrato: Tiempo trabajado en la empresa (fecha de entrada, fecha de salida)		
Actividad de la empresa: ¿A qué se dedica la empresa?		
Puesto de trabajo: ¿Qué puesto ocupa el trabajador en la empresa?		
Tareas: ¿Qué tipo de tareas desempeña el trabajador en su trabajo?		
Exposiciones: ¿El trabajador está expuesto en su trabajo a sustancias químicas (metales, disolventes, etc.), a agentes físicos (ruido, vibraciones, radiaciones, temperaturas extremas) o a agentes biológicos (bacterias, virus)? En caso afirmativo, especificar naturaleza y tipo de exposición.		

Periodo de contrato	Actividad de la empresa	Puesto de trabajo	Tareas	Exposiciones
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				

P. DATOS SOBRE EL PADRE

P1. ¿Qué edad tiene el padre?. Fecha de nacimiento
o años

P1
P1a

P2. ¿Cuál es su lugar de nacimiento?:

P2a. Municipio

P2b. Provincia/Región

P2c. País

.....

.....

.....

P2a P2b P2c

P3. ¿A que grupo étnico pertenece? (raza de su pareja)

- 1 Blanco
2 Asiático
3 Negro
4 Árabe (incluidos Norte de África y Oriente Medio)
5 Gitano

Nativo americano

Otros, especificar

9 Ns/Nc

P3

P4. ¿Cuál es la talla del padre aproximada? cm

P5. ¿Cuál es el peso del padre aproximado? Kg

P4

P5

P6

P6. ¿Qué nivel de estudios ha finalizado el padre? (no leer las opciones)

- 1 No sabe leer ni escribir
- 2 Sin estudios o estudios primarios incompletos
- 3 Estudios primarios (EGB, primaria, ESO)
- 4 Estudios secundarios (BUP, COU, FP)
- 5 Estudios universitarios
- 6 Otros (especificar).....

PL. HISTORIA LABORAL DEL PADRE

PL1. ¿Cuál es la situación laboral actual del padre?

- 1 Trabajador
- 2 Parado
- 3 Estudiante
- 4 Baja laboral Por:
- 5 Pensionista Por:.....
- 6 Otras (especificar):
- 9 Ns/Nc

SI LA RESPUESTA ES 1 ó 4 PASAR A LA PREGUNTA PL3

PL1

PL2. (Sólo si no está trabajando actualmente) ¿Ha trabajado el padre alguna vez durante los últimos 10 años?

- 1 Sí
- 2 No
- 9 Ns/Nc

}

SI LA RESPUESTA ES NO o Ns/Nc PASAR A LA PREGUNTA PC1

PL2

PL3. En relación con el trabajo actual y/o trabajos anteriores realizados por el padre durante los últimos 10 años, podría indicarnos (empezar por trabajo actual o último y continuar hacia atrás hasta cubrir los 10 años si ha habido otros trabajos previos)

Periodo de contrato: Tiempo trabajado en la empresa (fecha de entrada, fecha de salida)

Actividad de la empresa: ¿A qué se dedica la empresa?

Puesto de trabajo: ¿Qué puesto ocupa el trabajador en la empresa?

Tareas: ¿Qué tipo de tareas desempeña el trabajador en su trabajo?

Exposiciones: ¿El trabajador está expuesto en su trabajo a sustancias químicas (metales, disolventes, etc.), a agentes físicos (ruido, vibraciones, radiaciones, temperaturas extremas) o a agentes biológicos (bacterias, virus)? En caso afirmativo, especificar naturaleza y tipo de exposición.

Periodo de contrato	Actividad de la empresa	Puesto de trabajo	Tareas	Exposiciones
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				
Desde / Hasta / Mes Año				

PC. HISTORIA DE CONSUMO

PC1. Durante una semana normal, ¿cuántos vasos, copas o consumiciones acostumbra a tomar su pareja?

Entre semana Fin de semana

a. Cerveza

b. Vino o cava

c. Coñac, licores

d. Whisky, ginebra, vodka, ron

PC1ad PC1as

PC1bd PC1bs

PC1cd PC1cs

PC1dd PC1ds

R. HISTORIA REPRODUCTIVA**RP. HISTORIA OBSTÉTRICA PREVIA**

RP1. ¿Cuántos embarazos en total, a parte del actual, ha tenido Ud.?

SI LA RESPUESTA ES **NINGUNO** PASAR A LA PREGUNTA **RAC1**

RP2. ¿Notó un aumento del vello en la cara, en el pecho o en otras partes del cuerpo en algún embarazo anterior?

- 1 Sí
 2 No ☐
 9 Ns/Nc

RP2a. ¿En cuántos embarazos?

→ RP3. ¿Algún médico le ha diagnosticado alguna de las siguientes complicaciones en embarazos anteriores?

	Nº de orden del embarazo en el que tuvo el problema									
	1	2	3	4	5	6	7	8	9	10
a. Ningún problema										
b. Hipertensión en el embarazo										
c. Preeclampsia										
d. Eclampsia										
e. Diabetes gestacional										
f. Incompetencia cervical										
g. Otras, especificar										

RP1

RP2

RP2a

RP3a

RP3b

RP3c

RP3d

RP3e

RP3f

RP3g

A CONTINUACIÓN LE VOY A HACER UNAS PREGUNTAS RELACIONADAS CON SUS EMBARAZOS Y PARTOS ANTERIORES

(Por favor tenga en cuenta todos los embarazos no importa cual haya sido su término)

RE. HISTORIA EMBARAZOS ANTERIORES										
Nº	RE.1 Aborto 1) Espontáneo 2) Inducido 3) No	RE.2 Fecha de nacimiento o término del embarazo. (Día Mes Año)	RE.3 Recién nacido Vivo Muerto	RE.4. Duración embarazo (semanas)	RE.5 Peso en gramos del recién nacido	RE.6 Sexo Niño Niña 9) Ns/Nc	RE.7 Tipo de parto 1) Vaginal 2) Cesárea 3) Fórceps 4) Ventosa	RE.8 Semanas de lactancia	RE.9 Malformación en el bebé 1) Si 2) No	RE.10 ¿Vive en la actualidad? 1) Si 2) No
1	RE1.1 <input type="checkbox"/>	RE2.1 <input type="text"/>	RE3.1 <input type="checkbox"/>	RE4.1 <input type="text"/>	RE5.1 <input type="text"/>	RE6.1 <input type="checkbox"/>	RE7.1 <input type="checkbox"/>	RE8.1 <input type="text"/>	RE9.1 <input type="checkbox"/>	RE10.1 <input type="checkbox"/>
2	RE1.2 <input type="checkbox"/>	RE2.2 <input type="text"/>	RE3.2 <input type="checkbox"/>	RE4.2 <input type="text"/>	RE5.2 <input type="text"/>	RE6.2 <input type="checkbox"/>	RE7.2 <input type="checkbox"/>	RE8.2 <input type="text"/>	RE9.2 <input type="checkbox"/>	RE10.2 <input type="checkbox"/>
3	RE1.3 <input type="checkbox"/>	RE2.3 <input type="text"/>	RE3.3 <input type="checkbox"/>	RE4.3 <input type="text"/>	RE5.3 <input type="text"/>	RE6.3 <input type="checkbox"/>	RE7.3 <input type="checkbox"/>	RE8.3 <input type="text"/>	RE9.3 <input type="checkbox"/>	RE10.3 <input type="checkbox"/>
4	RE1.4 <input type="checkbox"/>	RE2.4 <input type="text"/>	RE3.4 <input type="checkbox"/>	RE4.4 <input type="text"/>	RE5.4 <input type="text"/>	RE6.4 <input type="checkbox"/>	RE7.4 <input type="checkbox"/>	RE8.4 <input type="text"/>	RE9.4 <input type="checkbox"/>	RE10.4 <input type="checkbox"/>
5	RE1.5 <input type="checkbox"/>	RE2.5 <input type="text"/>	RE3.5 <input type="checkbox"/>	RE4.5 <input type="text"/>	RE5.5 <input type="text"/>	RE6.5 <input type="checkbox"/>	RE7.5 <input type="checkbox"/>	RE8.5 <input type="text"/>	RE9.5 <input type="checkbox"/>	RE10.5 <input type="checkbox"/>
6	RE1.6 <input type="checkbox"/>	RE2.6 <input type="text"/>	RE3.6 <input type="checkbox"/>	RE4.6 <input type="text"/>	RE5.6 <input type="text"/>	RE6.6 <input type="checkbox"/>	RE7.6 <input type="checkbox"/>	RE8.6 <input type="text"/>	RE9.6 <input type="checkbox"/>	RE10.6 <input type="checkbox"/>
7	RE1.7 <input type="checkbox"/>	RE2.7 <input type="text"/>	RE3.7 <input type="checkbox"/>	RE4.7 <input type="text"/>	RE5.7 <input type="text"/>	RE6.7 <input type="checkbox"/>	RE7.7 <input type="checkbox"/>	RE8.7 <input type="text"/>	RE9.7 <input type="checkbox"/>	RE10.7 <input type="checkbox"/>
8	RE1.8 <input type="checkbox"/>	RE2.8 <input type="text"/>	RE3.8 <input type="checkbox"/>	RE4.8 <input type="text"/>	RE5.8 <input type="text"/>	RE6.8 <input type="checkbox"/>	RE7.8 <input type="checkbox"/>	RE8.8 <input type="text"/>	RE9.8 <input type="checkbox"/>	RE10.8 <input type="checkbox"/>
9	RE1.9 <input type="checkbox"/>	RE2.9 <input type="text"/>	RE3.9 <input type="checkbox"/>	RE4.9 <input type="text"/>	RE5.9 <input type="text"/>	RE6.9 <input type="checkbox"/>	RE7.9 <input type="checkbox"/>	RE8.9 <input type="text"/>	RE9.9 <input type="checkbox"/>	RE10.9 <input type="checkbox"/>
10	RE1.0 <input type="checkbox"/>	RE2.0 <input type="text"/>	RE3.0 <input type="checkbox"/>	RE4.0 <input type="text"/>	RE5.0 <input type="text"/>	RE6.0 <input type="checkbox"/>	RE7.0 <input type="checkbox"/>	RE8.0 <input type="text"/>	RE9.0 <input type="checkbox"/>	RE10.0 <input type="checkbox"/>

RE11. Si ha contestado “Sí” en alguna malformación del niño/a, especificar el tipo de malformación (si más de un niño, marcar nº de embarazo y especificar)

Tipo de Malformación/es _____

RE12. Si ha contestado “No” a vive en la actualidad, especificar (si más de un niño marcar n° de embarazo y especificar)

Edad (*en días, meses o años*) _____, **causa de la muerte** _____

RAC. ANTICONCEPTIVOS**RAC1. ¿Ha utilizado alguna vez el DIU como anticonceptivo?**

- 1 Sí
2 No
9 Ns/Nc

RAC1a. Si el DIU era hormonal, por favor indique durante cuanto tiempo. (calcular el tiempo para cada periodo en los que los utilizó, si ha habido interrupciones, y después sumar el total de meses y/o años):

- 1 < 1 año
2 1-2 años
3 2-5 años
4 > 5 años
9 Ns/Nc

RAC2. ¿Ha utilizado alguna vez anticonceptivos orales, inyecciones o implantes?

- 1 Sí
2 No
9 Ns/Nc

RAC2a. ¿Durante cuanto tiempo los ha utilizado? (calcular el tiempo para cada periodo en los que los utilizó, si ha habido interrupciones, y después sumar el total de meses y/o años)

- 1 < 1 año
2 1-2 años
3 3-5 años
4 6-10 años
5 11-15 años
>15 años
9 Ns/Nc

RAC3. ¿Ha utilizado usted algún método anticonceptivo antes de este embarazo?

- 1 Sí
2 No

SI LA RESPUESTA ES **NO** PASAR A LA PREGUNTA RA1**RAC3a. ¿Me podría decir cuál? (el último utilizado)**

	SÍ	NO	Ns/Nc	
1. Dispositivo, DIU	1	2	9	RAC3a1
2. Inyecciones	1	2	9	RAC3a2
3. Anticonceptivos orales	1	2	9	RAC3a3
4. Ovulos	1	2	9	RAC3a4
5. Esponja con espermicida	1	2	9	RAC3a5
6. Diafragma	1	2	9	RAC3a6
7. Preservativo o condón	1	2	9	RAC3a7
8. Retiro o coito interrumpido	1	2	9	RAC3a8
9. Otros, especificar:	1	2	9	RAC3a9

RAC3b. ¿Cuándo dejó de usar ese método anticonceptivo?.

Fecha:
Día Mes Año

RAC3c. ¿Lo dejó de usar para quedarse embarazada?

- 1 Sí
2 No
9 Ns/Nc

SI LA RESPUESTA ES **SI** PASAR A LA PREGUNTA RA3RAC1 RAC1a RAC 2 RAC2a RAC3 RAC3a1 RAC3a2 RAC3a3 RAC3a4 RAC3a5 RAC3a6 RAC3a7 RAC3a8 RAC3a9 RAC3b RAC3c

RA. EMBARAZO ACTUAL**RA1. Cuando se quedó embarazada, ¿tenía intención de quedarse embarazada?**

- 1 Sí
 No
 No se lo había planteado
 9 Ns/Nc

RA1 **RA2. En que fecha empezó a intentar quedar embarazada?**(mes, año): RA2 **RA3. ¿Cuántas visitas prenatales ha realizado en el Centro de Atención Primaria (CAP) antes de esta primera visita al hospital? Nº** RA3 **RA4. ¿Cuántas visitas prenatales ha realizado en la consulta de un ginecólogo privado antes de esta primera visita al hospital? Nº** RA4 **RA5. Fecha de la 1ª visita prenatal (CAP o ginecólogo privado):**RA5 **RA6. ¿Aproximadamente cuál era su peso antes de su actual embarazo?:** KgRA6 **RA7. Talla en cm (medirla):** cmRA7 **RA8. ¿Ha tenido usted vómitos en el presente embarazo?**

- 1 Sí
 2 No
 9 Ns/Nc

RA8 **RA8a. ¿En que mes del embarazo empezó?** RA8a **RA8b. ¿Por cuánto tiempo? (en semanas)** RA8b **RA8c. ¿Ha sido hospitalizada a causa de los mismos?**

- 1 Sí
 2 No
 9 Ns/Nc

RA8c **RA9. ¿Ha tenido usted algún episodio febril desde el comienzo del embarazo?**

- 1 Sí
 2 No
 9 Ns/Nc

RA9 **RA9a. ¿en qué semana del embarazo?** RA9a **RA9b. Fecha del episodio (mes, año)** RA9b **RA10. ¿Ha notado un aumento del vello en la cara, en el pecho o en otras partes del cuerpo en el actual embarazo?**

- 1 Sí
 2 No
 Ns/Nc

RA10 **RA10a. ¿Qué le diagnosticó el médico?**RA10a

M. ANTECEDENTES MEDICOS

MF. ENFERMEDADES FAMILIARES

MF1. ¿En los últimos *seis meses*, le han sido diagnosticadas por un médico enfermedades infecciosas a personas que conviven con usted (*tuberculosis, VIH, enfermedades exantemáticas víricas como rubéola, varicela, etc.*):

MF1a. Especificar tipo de problema		MF1b. Persona afectada	
	MF1a.1 <input type="checkbox"/>		MF1b.1 <input type="checkbox"/>
	MF1a.2 <input type="checkbox"/>		MF1b.2 <input type="checkbox"/>
	MF1a.3 <input type="checkbox"/>		MF1b.3 <input type="checkbox"/>
	MF1a.4 <input type="checkbox"/>		MF1b.4 <input type="checkbox"/>
	MF1a.5 <input type="checkbox"/>		MF1b.5 <input type="checkbox"/>
Entre su familia cercana y la del padre del bebé			
MF2. ¿Han existido problemas? :		Sí	No
			Ns/Nc
a) Para tener hijos	1	2	9
b) De abortos repetidos	1	2	9
c) De esterilidad (sin hijos)	1	2	9
Si todas las respuestas son No , pase a MF5			
MF3. Especificar tipo de problema		MF4. Familiar afectado	
	MF3.1 <input type="checkbox"/>		MF4.1 <input type="checkbox"/>
	MF3.2 <input type="checkbox"/>		MF4.2 <input type="checkbox"/>
	MF3.3 <input type="checkbox"/>		MF4.3 <input type="checkbox"/>
	MF3.4 <input type="checkbox"/>		MF4.4 <input type="checkbox"/>
	MF3.5 <input type="checkbox"/>		MF4.5 <input type="checkbox"/>
MF5. ¿Sabe si alguien de su familia, usted misma, el padre del bebé o la familia de éste, nacieron presentando?:		Sí	No
			Ns/Nc
a) Enfermedades hereditarias o cromosómicas	1	2	9
b) Malformaciones congénitas	1	2	9
c) Enfermedades graves	1	2	9
Si todas las respuestas son No , pase a MC1			
MF6. Especificar tipo de problema		MF7. Familiar Afectado	
	MF6.1 <input type="checkbox"/>		MF7.1 <input type="checkbox"/>
	MF6.2 <input type="checkbox"/>		MF7.2 <input type="checkbox"/>
	MF6.3 <input type="checkbox"/>		MF7.3 <input type="checkbox"/>
	MF6.4 <input type="checkbox"/>		MF7.4 <input type="checkbox"/>

MC. ENFERMEDADES CRÓNICAS DE LA MADRE**MC1. ¿Algún médico le ha diagnosticado alguna de las siguientes enfermedades?**

Enfermedades	Sí	No	Edad al diagnóstico
a. Diabetes (sin relación con los embarazos)	1	2	
b. Enfermedades cardíacas	1	2	
c. Alteraciones de la coagulación sanguínea	1	2	
d. Enfermedades renales y suprarrenales	1	2	
e. Alteraciones de la glándula tiroides	1	2	
f. Tuberculosis	1	2	
g. Infección por VIH	1	2	
h. Ansiedad	1	2	
i. Depresión	1	2	
j. Afección del tracto urinario	1	2	
k. Enfermedad intestinal inflamatoria crónica	1	2	
l. Tumores del aparato reproductor	1	2	
m. Otras, especificar:	1	2	

MC1a ☐ ☐ ☐

MC1b ☐ ☐ ☐

MC1c ☐ ☐ ☐

MC1d ☐ ☐ ☐

MC1e ☐ ☐ ☐

MC1f ☐ ☐ ☐

MC1g ☐ ☐ ☐

MC1h ☐ ☐ ☐

MC1i ☐ ☐ ☐

MC1j ☐ ☐ ☐

MC1k ☐ ☐ ☐

MC1l ☐ ☐ ☐

MC1m ☐ ☐ ☐

MM. HISTORIA MEDICA**MM1. ¿Le han realizado radiografías (rayos X) en los últimos 10 años?**

- 1 Sí
 2 No
 9 Ns/Nc



SI LA RESPUESTA ES **No** o **Ns/Nc**
 PASAR A LA PREGUNTA **MM1b**

MM1 1a ☐**MM1a. ¿Cuántas radiografías le han realizado, cuando, en qué parte del cuerpo y porqué en los últimos 10 años?**

Nº radiografías	Año	Parte del cuerpo	Motivo

MM1b. ¿Le realizaron radiografías antes de los 18 años?

- 1 Sí
 2 No
 9 Ns/Nc

MM1b ☐

<p>MM2. ¿Le han realizado alguna tomografía computerizada (TC o TAC) en los últimos 10 años?</p> <p>1 Sí</p> <p>2 No <input checked="" type="checkbox"/></p> <p>9 Ns/Nc</p> <p>MM2a. ¿cuántas TAC le han realizado, cuando, en qué parte del cuerpo y porqué en los últimos 10 años?</p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr> <th style="width: 15%;">Nº TAC</th> <th style="width: 15%;">Año</th> <th style="width: 35%;">Parte del cuerpo</th> <th style="width: 35%;">Motivo</th> </tr> </thead> <tbody> <tr><td> </td><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td><td> </td></tr> </tbody> </table> <p>MM2b. ¿Le realizaron un TAC antes de los 18 años?</p> <p>1 Sí</p> <p>2 No</p> <p>9 Ns/Nc</p> <p>MM3. ¿Ha recibido alguna vez una transfusión?</p> <p>1 Sí</p> <p>2 No <input checked="" type="checkbox"/></p> <p>9 Ns/Nc</p> <p>MM3a. Motivo (de la última).....</p> <p>MM3b. Fecha <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>→ MM4. ¿Ha visitado al dentista durante este embarazo?</p> <p>1 Sí</p> <p>2 No <input checked="" type="checkbox"/></p> <p>9 Ns/Nc</p> <p>MM4a. Motivo.....</p> <p>MM4b. Fecha <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>→ MM5. ¿Tiene empastes en la boca?</p> <p>1 Sí</p> <p>2 No <input checked="" type="checkbox"/></p> <p>9 Ns/Nc</p> <p>MM5a. Tipo 1: blanco, 2: metálico MM5b. Número <input type="text"/> <input type="text"/></p> <p>MM6. ¿Le ha sido realizado algún empaste durante el embarazo?</p> <p>1 Sí</p> <p>2 No <input checked="" type="checkbox"/></p> <p>9 Ns/Nc</p> <p>MM6a. Indique la fecha <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p>	Nº TAC	Año	Parte del cuerpo	Motivo																	<p>MM2 <input type="text"/></p> <p>MM2b <input type="text"/></p> <p>MM3 <input type="text"/></p> <p>MM3a <input type="text"/></p> <p>MM3b <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>MM4 <input type="text"/></p> <p>MM4a <input type="text"/></p> <p>MM4b <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>MM5 <input type="text"/></p> <p>MM5a <input type="text"/> MM5b <input type="text"/> <input type="text"/></p> <p>MM6 <input type="text"/></p> <p>MM6a <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p>
Nº TAC	Año	Parte del cuerpo	Motivo																		

→ MM7. ¿Le ha aplicado flúor el dentista durante el embarazo?

- | | | |
|---|-------|---|
| 1 | Sí | } |
| 2 | No | |
| 9 | Ns/Nc | |

SI LA RESPUESTA ES **No** o **Ns/Nc**
PASAR A LA PREGUNTA SIGUIENTE

MM7a. Indique la fecha

MM7

MM7a

A. ANTECEDENTES ALÉRGICOS								
	Madre			Padre			Hijos ¿En cuántos hijos?	
	Sí	No		Sí	No			
A1. Asma alérgico	1	2	A1a <input type="text"/>	1	2	A1b <input type="text"/>		A1c <input type="text"/>
A2. Dermatitis atópica	1	2	A2a <input type="text"/>	1	2	A2b <input type="text"/>		A2c <input type="text"/>
A3. Eccema	1	2	A3a <input type="text"/>	1	2	A3b <input type="text"/>		A3c <input type="text"/>
A4. Rinitis alérgica	1	2	A4a <input type="text"/>	1	2	A4b <input type="text"/>		A4c <input type="text"/>
A5. Otras, especificar	1	2	A5a <input type="text"/>	1	2	A5b <input type="text"/>		A5c <input type="text"/>

F. CONSUMO DE FARMACOS									
F1. ¿Ha tomado alguna medicación (de forma esporádica o habitual) desde un mes antes de quedarse embarazada y durante el presente embarazo? NOTA: INCLUIR los tratamientos adquiridos en herboristerías y/o automedicación, así como pomadas tales como cremas de corticoides. NO INCLUIR suplementos vitamínicos o minerales.									
Nombre del medicamento		Motivo/Enfermedad		Fecha inicio	Fecha fin	Frecuencia	Posología	Comentarios	¿Quién se lo prescribió?
Por ejemplo. Effergan	<input type="text"/>	Dolor muscular	<input type="text"/>						

V. VACUNAS		
V1. ¿Ha recibido alguna vacuna durante el último año y durante el presente embarazo?(viajes internacionales incluidos)		
Tipo de Vacuna	Fecha de administración (mes/año)	
	<input type="text"/>	<input type="text"/>
	<input type="text"/>	<input type="text"/>
	<input type="text"/>	<input type="text"/>
	<input type="text"/>	<input type="text"/>
	<input type="text"/>	<input type="text"/>
	<input type="text"/>	<input type="text"/>

C. COMPLICACIONES PRESENTADAS DESDE EL COMIENZO DEL EMBARAZO

Nota: En caso necesario añadir hojas suplementarias.

C1. ¿Ha presentado algún tipo de problema? Fiebre Infección de orina Tensión arterial alta Pérdida de líquido o sangre por la vagina Contracciones Otras, especificar	C1.1 ¿Le han atendido en un Centro de Atención Primaria a causa de ese problema? 1 Sí 2 No C1.1a ¿Dónde y quién le atendió? C1.1b Fecha: C1.1c ¿Cuál fue el diagnóstico? C1.1d ¿Tratamiento recibido? C1.1e ¿Información obtenida de la cartilla maternal? 1 Sí 2 No	C1.2. ¿Ha sido atendida en Urgencias o en un Centro Especializado o por un Especialista Privado a causa de ese problema? 1 Sí 2 No C1.2a. Especificar lugar y nombre del médico/a C1.2b. Fecha: C1.2c ¿Cuál fue el diagnóstico? C1.2d ¿Tratamiento recibido? C1.2e. ¿Información obtenida de un informe médico/a? 1 Sí 2 No
C2. ¿Ha presentado algún tipo de problema? Fiebre Infección de orina Tensión arterial alta Pérdida de líquido o sangre por la vagina Contracciones Otras, especificar	C2.1 ¿Le han atendido en un Centro de Atención Primaria a causa de ese problema? 1 Sí 2 No C2.1a ¿Dónde y quién le atendió?..... C2.1b Fecha: C2.1c ¿Cuál fue el diagnóstico? C2.1d ¿Tratamiento recibido? C2.1e ¿Información obtenida de la cartilla maternal? 1 Sí 2 No	C2.2. ¿Ha sido atendida en Urgencias o en un Centro Especializado o por un Especialista Privado a causa de ese problema? 1 Sí 2 No C2.2a. Especificar lugar y nombre del médico/a C2.2b. Fecha: C2.2c ¿Cuál fue el diagnóstico? C2.2d ¿Tratamiento recibido? C2.2e. ¿Información obtenida de un informe médico/a? 1 Sí 2 No
C3. ¿Ha presentado algún tipo de problema? Fiebre Infección de orina Tensión arterial alta Pérdida de líquido o sangre por la vagina Contracciones Otras, especificar.....	C3.1 ¿Le han atendido en un Centro de Atención Primaria a causa de ese problema? 1 Sí 2 No C3.1a ¿Dónde y quién le atendió? C3.1b Fecha: C3.1c ¿Cuál fue el diagnóstico? C3.1d ¿Tratamiento recibido? C3.1e ¿Información obtenida de la cartilla maternal? 1 Sí 2 No	C3.2 ¿Ha sido atendida en Urgencias o en un Centro Especializado o por un Especialista Privado a causa de ese problema? 1 Sí 2 No C3.2a. Especificar lugar y nombre del médico/a C3.2b. Fecha: C3.2c ¿Cuál fue el diagnóstico? C3.2d ¿Tratamiento recibido? C3.2e. ¿Información obtenida de un informe médico/a? 1 Sí 2 No

FINAL DE LA VISITA DE INCLUSIÓN (V1)

RECOGIDA DE MUESTRAS PARA INMA			
Hora finalización del Cuestionario General y CFA		_ _ : _ _	
CFA Nº 1 Realizado		Sí NO	
Si NO realizado CFA Nº 1 → Fecha Prevista		_ _ _ _ _ _ _ _	
Se le ha realizado el SCREENING de La Fe	Muestra Sangre ECO	SI	NO
Si no, fecha prevista del SCREENING	Muestra Sangre ECO	SI	NO
		_ _ _ _	_ _ _ _
Fecha de la próxima cita del hospital "LA FE" (Semana 20)		_ _ _ _ _ _ _ _	
Recogida orina de la semana 12		SI	NO
Recogida uña en la semana 12		SI	NO
COMPRESION:			
Nivel estimado de comprensión del cuestionario:		1. Excelente 2. Buena 3. Regular 4. Mala	
¿Quién responde a las preguntas referidas al marido o compañero?		1. El mismo 2. La mujer 3. Otros.....	
COMENTARIOS DEL ENTREVISTADOR			

12.2. Annexe 2: Sintaxi en R de la simulació del bootstrap

```

library(boot)

#####
# Case with calibration model having R2~72%
# and NO2 being associated with weight
#####

# x1: covariate to predict no2

set.seed(8234259)
x1=rnorm(50)
no2=1.5+2.6*x1 +rnorm(50,0,1.4)
data.no2=data.frame(no2,x1)

x1.weight=rnorm(300,0)
no2.unobserved=1.5+2.6*x1.weight +rnorm(50,0,1.8)
weight=1+2*no2.unobserved+rnorm(300,0,1.5)
# so, true value of coefficient = 2
summary(lm(weight~no2.unobserved))

data.weight=data.frame(weight,x1=x1.weight)
calib1 <- lm(no2~x1,data=data.no2)
summary(calib1) # R2=72%
data.weight$Xcal1 <- predict(calib1, newdata=data.weight)
summary(lm(weight~Xcal1,data=data.weight))

fit.model = function(data.weight,data.no2) {
  calib1 <- lm(no2~x1,data=data.no2)
  data.weight$Xcal1 <- predict(calib1, newdata=data.weight)
  mod.calib1 <- lm(weight~Xcal1,data=data.weight)
  summary(mod.calib1)$coefficients[2,1]
}

calib.fun=function(data.weight,i) {
  j=sample(1:dim(data.no2)[1],dim(data.no2)[1],replace=T)
  fit.model(data.weight[i,],data.no2[j,])
}
calib.boot=boot(data.weight,calib.fun,R=500)
summary(lm(weight~Xcal1,data=data.weight))
boot.ci(calib.boot)

#####
# Case with calibration model having R2~72%
# and NO2 NOT being associated with weight

```

```
#####
```

```
set.seed(8234259)
x1=rnorm(50)
no2=2.43+2.6*x1 +rnorm(50,0,1.3)
data.no2=data.frame(no2,x1)
x1.weight=rnorm(300,0)
no2.unobserved=1.5+2.6*x1.weight +rnorm(50,0,1.8)
x3=rnorm(300)
weight=1+2*x3+0.05*no2.unobserved+rnorm(300,0,4)
data.weight=data.frame(weight,x1=x1.weight,x3)
summary(lm(no2~x1,data=data.no2))
calib1 <- lm(no2~x1,data=data.no2)
data.weight$Xcal1 <- predict(calib1, newdata=data.weight)
summary(lm(weight~Xcal1+x3,data=data.weight))
# now weight associated to x3 but not with no2

fit.model = function(data.weight,data.no2) {
  calib1 <- lm(no2~x1,data=data.no2)
  data.weight$Xcal1 <- predict(calib1, newdata=data.weight)
  mod.calib1 <- lm(weight~x3+Xcal1,data=data.weight)
  summary(mod.calib1)$coefficients[3,1]
}

calib.fun=function(data.weight,i) {
  j=sample(1:dim(data.no2)[1],dim(data.no2)[1],replace=T)
  fit.model(data.weight[i,],data.no2[j,])
}
calib.boot=boot(data.weight,calib.fun,R=500)
boot.ci(calib.boot)
summary(lm(weight~no2.unobserved,data=data.weight))
```

```
#####
##
```

```
# simulation to check that the coverage is 95%, i.e. 95% of the times, the
# interval includes the true value
```

```
#####
###
```

```
simulate.dataset1=function() {
  x1=rnorm(50)
  no2=1.5+2.6*x1 +rnorm(50,0,1.4)
  data.no2=data.frame(no2,x1)
  data.no2
}
```

```
simulate.dataset2=function() {
  x1.weight=rnorm(300,0)
  no2.unobserved=1.5+2.6*x1.weight +rnorm(50,0,1.8)
```

```

weight=1+2*no2.unobserved+rnorm(300,0,1.5)
# so, true value of coefficient = 2
data.weight=data.frame(weight,x1=x1.weight)
data.weight
}

fit.model = function(data.weight,data.no2) {
  calib1 <- lm(no2~x1,data=data.no2)
  data.weight$Xcal1 <- predict(calib1, newdata=data.weight)
  mod.calib1 <- lm(weight~Xcal1,data=data.weight)
  summary(mod.calib1)$coefficients[2,1]
}

calib.fun=function(data.weight,i) {
  j=sample(1:dim(data.no2)[1],dim(data.no2)[1],replace=T)
  fit.model(data.weight[i,],data.no2[j,])
}

fit.model.ci = function(data.weight,data.no2) {
  calib1 <- lm(no2~x1,data=data.no2)
  data.weight$Xcal1 <- predict(calib1, newdata=data.weight)
  mod.calib1 <- lm(weight~Xcal1,data=data.weight)
  c(summary(mod.calib1)$coefficients[2,1]-1.96*summary(mod.calib1)$coefficients[2,2],
    summary(mod.calib1)$coefficients[2,1]+1.96*summary(mod.calib1)$coefficients[2,2])
}

nsim=200
set.seed(8234259)

ci.lm=matrix(NA,nrow=nsim,ncol=2)
ci.boot=matrix(NA,nrow=nsim,ncol=2)
for (i in 1:nsim) {
  data.no2=simulate.dataset1()
  data.weight=simulate.dataset2()
  ci.lm[i,]=fit.model.ci(data.weight,data.no2)
  calib.boot=boot(data.weight,calib.fun,R=300)
  ci.boot[i,]=boot.ci(calib.boot)$percent[4:5]
}

coverage=function(true.value,ci) {
  # count how many intervals contain the true value
  100*sum( as.numeric(true.value>=ci[,1] & true.value<=ci[,2]))/dim(ci)[1]
}

coverage(2,ci.lm)
coverage(2,ci.boot)

# linear model -> confidence intervals too narrow -> underestimate variance
# bootstrap -> correct confidence interval

```

12.3. Annexa 3: Variables de Trànsit

Variable	N	Mitjana	SD	Mediana	Min	Max
Distància al pàrking més proper (m)	57	273.7	285.3	182.9	36.3	1458.1
Distància al taller més proper (m)	57	207.4	222.2	151.6	14.1	1363.6
Distància a la benzinera més propera (m)	57	713.8	456.0	602.2	47.4	2392.3
Distància a la via més propera amb alta densitat de trànsit (m)	57	293.4	262.1	261.8	5.1	1067.4
Distància a la via més propera amb mitjana densitat de trànsit (m)	57	378.8	339.7	297.8	3.4	1459.4
Distància ponderada a la via més propera amb alta/mitjana densitat de trànsit (m)	57	265.7	336.4	179.5	5.1	1731.2
Distància a la via més propera amb alta/mitjana densitat de trànsit (m)	57	197.1	214.4	143.9	3.4	840.4
Km de vies d'alt transit en un radi de 50 m	57	0.0	0.1	0.0	0.0	0.4
Km de vies d'alt transit en un radi de 500 m	57	1.7	1.3	1.6	0.0	4.4
Km de vies d'alt transit en un radi de entre 50-150 m	57	0.1	0.2	0.0	0.0	1.0
Km de vies d'alt transit en un radi de entre 50-250 m	57	0.4	0.5	0.0	0.0	1.9
Km de vies de mitjà transit en un radi de 150 m	57	0.1	0.2	0.0	0.0	0.8
Km de vies de baix transit en un radi de 50 m	57	0.1	0.1	0.1	0.0	0.3
Km de vies de baix transit en un radi de entre 250_500 m	56	9.0	2.7	9.8	1.0	13.3
Km de vies de mitjà/alt transit en un radi de 50 m	57	0.1	0.1	0.0	0.0	0.5
Edificis en un radi de 50 metres	57	19.3	14.2	16.0	1.0	53.0
Edificis en un radi de 100 metres	57	72.9	48.8	68.0	2.0	192.0
Població en un radi de 50 m.	57	204.9	187.0	154.0	0.0	750.0
Hectàrees de terreny urbà i industrial en un radi de 200 m	57	11.1	3.3	12.6	0.0	18.9
Hectàrees de terreny urbà en un radi de 200 m	57	10.7	3.3	12.6	0.0	12.6